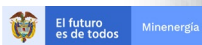
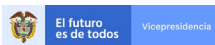


Manual de Referencia

en Mejores Prácticas de Gestión de Datos Oceánicos

Arqueología y recuperación de datos e información

2019



Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos Número 4 / 2019

Una publicación digital del Comité Técnico Nacional de Coordinación de Datos e Información Oceánica - CTN Diocean, de la Comisión Colombiana del Océano - CCO Bogotá D.C., Colombia www.cco.gov.co

DIRECCIÓN

Vicealmirante Juan Manuel Soltau Ospina
Director General Marítimo - DIMAR

Capitán de Navío Juan Camilo Forero Hauzeur
Secretario Ejecutivo de la CCO

CONTENIDOS

Darío Fernando León Rojas
Ministerio de Defensa Nacional - MDN

Ruby Viviana Ortiz Martínez
Wilder Antonio Álvarez Sánchez
Dirección General Marítima – Dimar

Martín Hernando González Martínez
Nancy Liliana Barreto Parra
John Guibsson García Guerrero
Fredy Alberto Gutiérrez García
Ana Alexandra Morales Escobar
Instituto Geográfico Agustín Codazzi – IGAC

Adriana Margarita Roza Melo
Parques Nacionales Naturales de Colombia – PNN

Pedro Augusto Rangel Segura
Rigoberto Blandón Grajales
Omar Gonzalo Santafé Alfonso
Jaime Alberto Garzón Barrios
Servicio Geológico Colombiano – SGC

REVISIÓN DE ESTILO Y CONTENIDO

Ruby Viviana Ortiz Martínez
Dirección General Marítima – Dimar

Carolina García
Instituto de Investigaciones Marinas y Costeras – Invemar

Néstor Ricardo Bernal Suárez
Universidad Distrital Francisco José de Caldas

CONCEPTO GRÁFICO

Área Comunicaciones Estratégicas (Acoes) - Dimar

EDITORIAL DIMAR

Fotografía:

Toma de muestras con roseta. Crucero Oceanográfico cuenca Pacífica colombiana LIV / XVIII Crucero Regional Conjunto CPPS, enero de 2016. Cortesía Dimar

ISSN 2539-2212 (en línea)



Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4 / 2019 por CCO se encuentra bajo una [Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos

Número 4 / 2019

Bogotá D.C., Colombia. Diciembre de 2019

El "Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos" es un producto del Grupo de Trabajo en Mejores Prácticas de Gestión de Datos (GT MPGD), el cual hace parte de la estructura interna del Comité Técnico Nacional de Coordinación de Datos e Información Oceánica (CTN Diocean) de la Comisión Colombiana del Océano (CCO), y tiene por objeto recomendar estándares y mejores prácticas para la gestión de datos e información oceánica del país en todo su ciclo de vida, teniendo en cuenta los niveles nacional, regional y local.

Es de carácter técnico e informativo, en idioma español y en formato electrónico. La información y conceptos expresados en esta publicación deben ser utilizados por los interesados bajo su responsabilidad y criterio, sin embargo, se entiende que cualquier divergencia con lo publicado es de interés del CTN Diocean, por lo que se agradece el envío de sus observaciones o sugerencias al correo electrónico oceano@cco.gov.co.

Este producto intelectual está protegido por el Copyright y cuenta con una política de acceso abierto para su consulta. Sus condiciones de reconocimiento, uso y distribución están definidas por el licenciamiento *Creative Commons* (CC), que expresa de antemano los derechos definidos.

Cítese la obra completa de la siguiente manera:

CTN Diocean. (2019). *Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4 / 2019*. Bogotá D.C., Colombia: DIMAR. DOI <https://doi.org/10.26640/25392212.4.2019>.

Cítense los capítulos así:

León Rojas, Darío Fernando. (2019). *Importancia de los metadatos en la recuperación de información*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp. 07-15). Bogotá, D.C., Colombia: DIMAR.

Ortiz Martínez, Ruby Viviana. (2019). *Arqueología y recuperación de datos recopilados durante más de 40 años a bordo de cruceros oceanográficos colombianos*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp. 16-26). Bogotá, D.C., Colombia: DIMAR.

Álvarez Sánchez, Wilder Antonio. (2019). *Desafíos en la arqueología y recuperación de datos de la red de monitoreo meteomarinero más grande de Colombia*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp. 27-36). Bogotá, D.C., Colombia: DIMAR.

González Martínez, Martín Hernando; Barreto Parra, Nancy Liliana; García Guerrero, John Guibsson; Gutiérrez García, Fredy Alberto y Morales Escobar, Ana Alexandra. (2019). *Proceso de preservación de la colección de aerofotografías históricas de Colombia: inventario, digitalización y disposición para consulta y desarrollo de estudios aplicados*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp.37-44). Bogotá, D.C., Colombia: DIMAR.

Rangel Segura, Pedro Augusto y Blandón Grajales, Rigoberto. (2019). *Reseña de la arqueología y recuperación de información para el "Banco de Información Petrolera"*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp.45-48). Bogotá, D.C., Colombia: DIMAR.

Rozo Melo, Adriana Margarita. (2019). *Arqueología y recuperación de registros biológicos de Parques Nacionales Naturales de Colombia*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp.49-51). Bogotá, D.C., Colombia: DIMAR.

Santafé Alfonso, Omar Gonzalo y Garzón Barrios, Jaime Alberto. (2019). *Arqueología y recuperación de información para el mayor inventario geocientífico de Colombia*. (Ed.), Manual de Referencia en Mejores Prácticas de Gestión de Datos Oceánicos. Número 4/2019 (pp.52-55). Bogotá, D.C., Colombia: DIMAR.

CONTENIDO

Pág.

INTRODUCCIÓN.....	5
1. IMPORTANCIA DE LOS METADATOS EN LA RECUPERACIÓN DE INFORMACIÓN.....	7
1.1. ANTECEDENTES.....	7
1.2. ¿QUÉ SON LOS METADATOS?.....	8
1.3. CLASIFICACIÓN DE METADATOS.....	9
1.4. RECUPERACIÓN DE INFORMACIÓN.....	10
1.4.1. Semántica.....	11
1.4.2. Núcleo de metadatos.....	11
1.4.3. Operadores booleanos.....	13
1.4.4. Truncadores.....	14
1.5. REFLEXIONES FINALES.....	15
2. EXPERIENCIAS.....	16
2.1. ARQUEOLOGÍA Y RECUPERACIÓN DE DATOS RECOPIADOS DURANTE MÁS DE 40 AÑOS A BORDO DE CRUCEROS OCEANOGRÁFICOS COLOMBIANOS.....	16
2.1.1. Creación del Centro Colombiano de Datos Oceanográficos.....	16
2.1.2. La primera experiencia en arqueología y recuperación de datos.....	17
2.1.3. Implementación de “OceanData”.....	18
2.1.4. Las iniciativas “Cenimarc” y “Retrocean”.....	19
2.1.5. El Atlas de los Datos Oceanográficos de Colombia.....	21
2.1.6. Posicionamiento estratégico del Cecoldo.....	22
2.1.7. Conclusiones.....	26
2.2. DESAFÍOS EN LA ARQUEOLOGÍA Y RECUPERACIÓN DE DATOS DE LA RED DE MONITOREO METEOMARINO MÁS GRANDE DE COLOMBIA.....	27
2.2.1. Etapa 1. Búsqueda e inventario de datos.....	28
2.2.2. Etapa 2. Evaluación de las series de tiempo.....	28
2.2.3. Etapa 3. Estandarización de las mediciones.....	32
2.2.4. Etapa 4. Estructuración de la base de datos primaria.....	32
2.2.5. Etapa 5. Control de calidad de datos.....	34
2.2.6. Reflexiones finales.....	36
2.3. PROCESO DE PRESERVACIÓN DE LA COLECCIÓN DE AEROFOTOGRAFÍAS HISTÓRICAS DE COLOMBIA: INVENTARIO, DIGITALIZACIÓN Y DISPOSICIÓN PARA CONSULTA Y DESARROLLO DE ESTUDIOS APLICADOS.....	37
2.3.1. Preparación y almacenamiento.....	38
2.3.2. Manipulación.....	41

2.3.3.	Inventario	42
2.3.4.	Digitalización	42
2.3.5.	Documentación	43
2.3.6.	Disposición y uso	43
2.3.7.	Reflexiones finales	44
2.4.	RESEÑA DE LA ARQUEOLOGÍA Y RECUPERACIÓN DE INFORMACIÓN PARA EL “BANCO DE INFORMACIÓN PETROLERA”	45
2.4.1.	El primer “Banco de datos”	45
2.4.2.	Desarrollo de una base de datos	46
2.4.3.	El BIP hoy.....	47
3.	PERSPECTIVAS.....	49
3.1.	ARQUEOLOGÍA Y RECUPERACIÓN DE REGISTROS BIOLÓGICOS DE PARQUES NACIONALES NATURALES DE COLOMBIA	49
3.1.1.	Generalidades	49
3.1.2.	Avances alcanzados.....	50
3.2.	ARQUEOLOGÍA Y RECUPERACIÓN DE INFORMACIÓN PARA EL MAYOR INVENTARIO GEOCIENTÍFICO DE COLOMBIA	52
3.2.1.	Generalidades	52
3.2.2.	Avances y metas.....	53
4.	BIBLIOGRAFÍA	56
5.	SIGLAS.....	59

LISTA DE FIGURAS ►

Pág.

Figura 1. Plano de la biblioteca de la Abadía Benedictina, la cual es el centro del argumento en que se desarrolla la novela "El nombre de la Rosa".....	8
Figura 2. Ilustración del uso de los operadores booleanos.....	14
Figura 3. Portada del Informe de Crucero Oceanográfico “Océano IV (Áreas 2 y 3) – 1974” publicado en 1982.....	17
Figura 4. Aspecto de uno de los módulos de “OceanData” del CIOH.....	18
Figura 5. Metodología empleada en la fase I del proyecto “Retrocean”.....	20
Figura 6. Diagrama de flujo seguido en el procesamiento de datos oceanográficos.....	21
Figura 7. Fases de la metodología archivística para la recuperación de información histórico-oceanográfica.....	22
Figura 8. Imágenes del proceso de la arqueología de datos oceanográficos en el Cioh.....	23
Figura 9. Imágenes del proceso de la arqueología de datos oceanográficos en el Cccp.....	23
Figura 10. Cobertura geográfica de los conjuntos de datos oceanográficos y de meteorología marina 1969-2017 recuperados en el proyecto “Posicionamiento estratégico del Cecoldo”.....	25
Figura 11. Etapas del proceso de arqueología de datos meteorológicos y del nivel del mar de la RedMpomm.....	27
Figura 12. Datos fuera de rango de tiempo en una serie de velocidad del viento para estación de Buenaventura.....	29
Figura 13. En rojo, datos estimados en series de estaciones ubicadas en Buenaventura y Cartagena.	30
Figura 14. Presión barométrica de estación Buenaventura: (a) detalle del comportamiento o variabilidad esperados; (b) periodo de tiempo de análisis; y (c) detalle de datos imposibles.....	31
Figura 15. Resultado del proceso de calificación de datos de meteorología marina de la RedMpomm 2005-2018.....	35
Figura 16. Foto del cartel de la entrada de la bodega de rollos.....	39
Figura 17. Foto interna de la bodega (Planoteca).....	39
Figura 18. (a) Foto interna de la bodega con los estantes de rollos y (b) rollo fotográfico.....	40
Figura 19. Banda transportadora de rollos en la bodega.....	41
Figura 20. Manipulación de rollos.....	42
Figura 21. Sala de digitalización con escáneres automáticos.....	43
Figura 22. Tipo de medios desde donde se recuperaron datos para el BIP.....	45
Figura 23. Panorámica de la Cintoteca ubicada en el departamento de Cundinamarca.....	46
Figura 24. Transformación de medios de almacenamiento de datos y documentos técnicos E&P.....	47
Figura 25. Carpetas de información en SULA.....	50
Figura 26. Aspecto de la plataforma MIIG para la búsqueda de información geocientífica del país.....	53
Figura 27. Documento producido en 1929 escaneado con funcionalidad OCR para facilitar su recuperación y mapa anexo disponibles en formato PDF.....	54

LISTA DE TABLAS

Pág.

Tabla 1. Elementos del estándar internacional Dublin Core.	12
Tabla 2. Cantidad de cruceros cuyos datos oceanográficos fueron recuperados en el marco del proyecto “Posicionamiento estratégico del Cecoldo”.	24
Tabla 3. Agregación temporal de los sensores de datos de la RedMpomm.	32
Tabla 4. Datos meteorológicos y de nivel del mar de la RedMpomm recuperados a 31 de diciembre de 2018.....	33
Tabla 5. Extracto de banderas de calidad IODE aplicados por la RedMpomm.	34
Tabla 6. Grupos biológicos integrados en la temática de investigaciones de SULA.....	50

INTRODUCCIÓN

El Comité Técnico Nacional de Coordinación de Datos e Información Oceánica (CTN Diocean) se permite presentar el cuarto número de su “Manual de Referencia en Mejores Prácticas de Datos Oceánicos”, que aborda en esta oportunidad la “arqueología y recuperación” de datos e información. Esta temática, en el contexto de la gestión de datos y más específicamente de la iniciativa *Global Oceanographic Data Archaeology and Rescue* (GODAR), hace referencia a un proceso de dos etapas: la primera “*archaeology*”, que consiste en buscar, identificar, evaluar y describir datos históricos para que sean útiles a la comunidad; y la segunda “*rescue*”, se refiere al esfuerzo de almacenar y conservar los datos, mediante la digitalización, copia en medios electrónicos, y archivo en bases de datos o repositorios digitales (Adaptado de IOC, 1999).

A través de esta publicación las instituciones y universidades miembros del CTN Diocean, dan a conocer el esfuerzo que viene realizando Colombia en la materia de “arqueología y recuperación”, para aumentar los archivos históricos y recientes de datos e información oceánica en formato digital, con la finalidad de disponerlos a la comunidad de manera estandarizada e interoperable para su reutilización.

En el primer capítulo se plantean reflexiones acerca de la importancia de los metadatos en la recuperación de la información, vista desde la perspectiva de catalogación, es decir, de la estructuración del dato a partir de documentar la mayor cantidad de información asociada a este, pero también desde el punto de vista de los aportes que realiza el metadato en los procesos de búsqueda de datos e información en internet.

En el segundo capítulo se reúnen experiencias en arqueología y recuperación de datos e información en diferentes disciplinas; las primeras de ellas corresponden a dos iniciativas lideradas por la Dirección General Marítima (Dimar) a través de las cuales se ha recuperado de una parte, el patrimonio histórico de datos oceanográficos y de meteorología marina obtenidos en los más de un centenar de cruceros oceanográficos desarrollados en Colombia; de otra parte, las series de tiempo de datos metoceanicos cuyas mediciones en tiempo real en diferentes puntos de la costa Caribe, Pacífico y áreas insulares colombianas, iniciaron en el año 2005 y se extienden a la actualidad.

El Instituto Geográfico Agustín Codazzi (IGAC) explica la complejidad del proceso de preparación, manipulación, inventario, almacenamiento, digitalización y documentación de la colección de aerofotografías históricas de Colombia, el cual ha contado con la asesoría técnica del Archivo General de la Nación (AGN), y cómo este esfuerzo ha permitido disponer parte del material en formato digital para consulta del ciudadano.

Por su parte, el Servicio Geológico Colombiano (SGC) hace un rápido repaso a la historia sobre cómo se ha ido estructurando el Banco de Información Petrolera (BIP) en los últimos años, en relación con las actividades desarrolladas para recuperar y llevar la información técnica de Exploración y Producción (E&P) a una organización tal, que permite disponer de manera fácil y segura para ser utilizada en el desarrollo de proyectos de interés nacional e internacional.

Por último, el tercer capítulo presenta los avances que tanto Parques Nacionales Naturales de Colombia (PNN) como el SGC han obtenido, adelantando actividades de arqueología y recuperación de datos e información, una vez han entrado en operación plataformas y software institucionales en disciplinas tales como las geociencias y la biología. Lo anterior demuestra que las mencionadas actividades están presentes en cualquiera de las etapas de los sistemas de información y con fines comunes, que son disponer y preservar datos e información histórica y reciente con la mejor calidad posible.

1. IMPORTANCIA DE LOS METADATOS EN LA RECUPERACIÓN DE INFORMACIÓN ►

1.1. Antecedentes

Todas las civilizaciones han producido datos, información y conocimiento; para evitar que estos caigan en el olvido de la oralidad, deben ser encapsulados en objetos que puedan ser almacenados y consultados; dichos objetos pueden ser los documentos, tal como lo expuso a principios del siglo XX la bibliotecóloga francesa Suzanne Briet “Todos los objetos pueden o no convertirse en documento [...] ejemplo un Antílope corriendo en las llanuras salvajes del África no es un documento, pero si es capturado y llevado a un zoológico o convertido en objeto de estudio sí lo es”.

En la actualidad, los documentos están disponibles en formatos físicos (generalmente en papel) o en formatos digitales que se consultan a través de medios electrónicos, y a medida que crece la generación de documentos es mayor el reto de administrarlos. Es así que para lograr difundir el conocimiento escrito deben crearse mecanismos que logren su fácil ubicación ya sea de forma física o digital, razón por la cual, desde los imperios antiguos hasta la época actual, la sociedad ha desarrollado bibliotecas o sistemas de información para administrar los documentos bajo procedimientos estandarizados de captura, control y difusión.

El primer paso para facilitar o restringir el acceso a los documentos, es definir el procedimiento de organización y distribución de la documentación, ya sea en un medio físico mediante estanterías o en un medio digital a través de una pantalla. Humberto Eco (1984) en su libro “El nombre de la Rosa” desarrolla una trama detectivesca en torno a un libro de Platón, el cual estaba sigilosamente custodiado en la biblioteca de una Abadía medieval, con organización en forma de laberinto y además con medios de seguridad como manchar con veneno las hojas de los libros, para evitar lo que actualmente se podría acercarse a una “fuga de Información”.

Para que esta organización permitiera que las personas que sí tenían acceso al libro de Platón logran la consulta sin mayor dificultad, el bibliotecario contaba con un documento en el cual en orden alfabético organizaba los títulos de cada libro almacenado en la biblioteca, y al lado de cada libro registraba una sigla y un número que representaban respectivamente el salón y el estante en donde este se ubicaba (Figura 1). El mencionado documento creado y usado por cada biblioteca del mundo durante más de mil años se le conoció como “Catálogo” y la información que contenía es llamada en el presente como “Metadatos”. En la actualidad gracias a las nuevas tecnologías de la información, los metadatos permiten direccionar hacia la ubicación de la información de forma cada vez más eficiente.

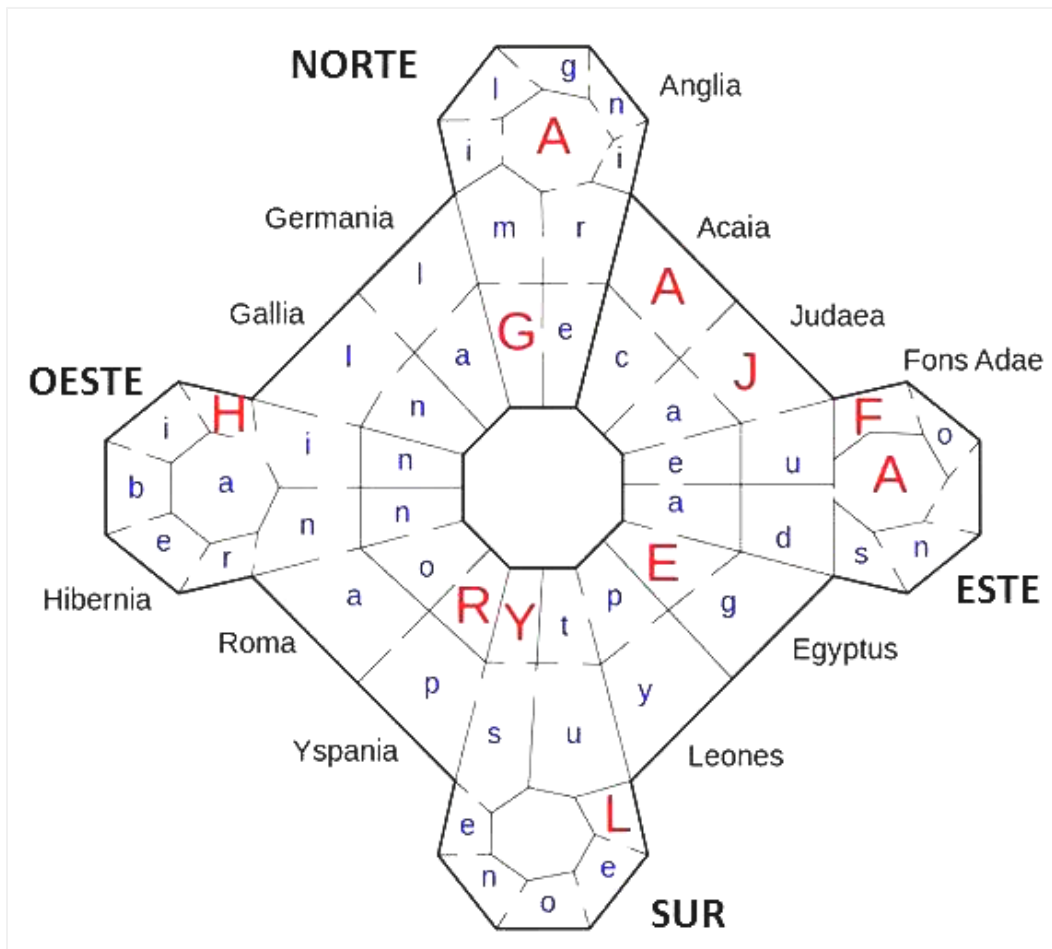


Figura 1. Plano de la biblioteca de la Abadía Benedictina, la cual es el centro del argumento en que se desarrolla la novela "El nombre de la Rosa"¹.

1.2. ¿Qué son los metadatos?

Sin importar si su aplicación es básica o avanzada, los metadatos cumplen la función de describir la información que, visto desde la perspectiva de Méndez Rodríguez (2003), se definen como:

- Datos altamente estructurados que describen la información, el contenido, la calidad, la condición y otras características de los datos.
- "Información sobre información" o "datos sobre los datos".

¹ Fuente: Eco, H. (1984) El Nombre de la Rosa. Barcelona: Lumen.

Esta última fue la definición inicial descrita por Richard Myers (Citado por Méndez Rodríguez, 2003) en la década de los sesenta del siglo XX, en donde propone que todo aquel componente que identifique a un documento es considerado como metadato. Otra concepción como la expuesta por NISO (2019), define los metadatos como “una estructura de información que describe, explica, localiza o de otro modo; hace la recuperación de la información lo más fácil posible para su uso y manejo”. Para su completa utilización Bearman y Sochats (Citado por Taylor, 2009), definen que los metadatos deben cumplir con ciertas condiciones de organización básicas las cuales son: Registro, términos y condiciones (de acceso), estructura, contexto, contenido e historia del uso del recurso.

En general, cualquier forma estructurada usada para describir y organizar información puede ser considerada como metadato, siempre y cuando cumplan con la función de direccionar a una información específica, que generalmente se encuentra al interior de un documento. Dentro de la bibliotecología el metadato es la base de su sistema de control y difusión bibliográfica; esta última en un principio se efectuaba por medio de formatos físicos como los ficheros o fichas de Descripción Bibliográfica Internacional Normalizada (ISBD), pero evolucionó gracias a la tecnología, hacia el almacenamiento en bases de datos bibliográficas, las cuales por medio de buscadores permiten al ser humano mayor eficiencia y precisión en la ubicación de la información, ya sea proporcionando puntos de acceso físicos o digitales.

Así mismo, en esta nueva era tecnológica los metadatos permiten la interoperabilidad de los sistemas de información, con ayuda de protocolos de transmisión tales como el Z39.50, haciendo visible información de manera automática proveniente de diferentes fuentes y separadas por grandes distancias, mejorando los tiempos de divulgación y comunicación del conocimiento; esto es como una “gran autopista [electrónica] de la información” (Méndez Rodríguez, 2002) normalizada.

1.3. Clasificación de metadatos

Debido a la amplia variedad de características y usos de la información, Méndez Rodríguez (2002) explica que existen metadatos para:

- Describir recursos (objetos) de información, como Dublin Core.
- Definir los elementos de los registros catalográficos en bibliotecas, dentro del cual se encuentra el formato *Machine Readable Cataloging* (MARC).
- El intercambio de información de museos, como el modelo *Consortium for the Interchange of Museum Information* (CIMI).
- Describir recursos geográficos o geoespaciales, como por ejemplo el estándar ISO 19115.
- Describir información electrónica gubernamental, por ejemplo, el servicio *Government Information Locator Service* (GILS).
- Para la descripción archivística, como puede ser el modelo *Encoded Archival Description* (EAD).

Por lo anterior, varios autores simplifican la clasificación de metadatos en los siguientes tipos, que en general, dependen de las necesidades de los usuarios internos y externos:

- **Administrativos:** Son metadatos que facilitan la gestión y administración de los objetos o documentos. Estos elementos se utilizan para el control y el mantenimiento de la calidad del objeto, definen la ubicación y el estado actual del mismo.
- **Descriptivos:** Son los componentes que identifican el recurso. Se basa en los elementos básicos de descripción (p. ej. Autor, título, editorial, etc.) más los elementos especiales que sean necesarios para representar el objeto.
- **Conservación:** Son los que se encargan de la conservación de la fuente de información como objeto de referencia. Estos elementos se hacen necesarios debido a la constante alteración que sufre la información.
- **Técnicos:** Estos metadatos se encargan de guardar información sobre las características y necesidades que posee el objeto para su visualización.
- **Uso:** Son metadatos que se generan automáticamente mediante un código que calcula el uso del objeto. Generalmente son los contadores que informan sobre el número de entradas o descargas de un recurso.

1.4. Recuperación de información

La “Web semántica” es un conjunto de herramientas tecnológicas de localización en internet las cuales, mediante la interacción con diferentes bases de datos de bibliotecas, repositorios digitales, sistemas de información, etc., buscan un camino adecuado y normalizado para la recuperación de la información y cubrir así una necesidad específica. Para ello se requieren datos con contenido, significado y preferiblemente relacionados con otros datos, evitando de esta manera sucesos como el del famoso cuento de Lewis Carroll “Alicia en el País de las Maravillas”, en donde Alicia le pregunta al gato de Cheshire en un cierto punto qué camino debía tomar, y Cheshire le contesta: “Eso depende mucho del lugar adonde quieras ir. Si no sabes a dónde quieres ir, no importa qué camino sigas, cualquier camino te llevará a alguna parte”.

Los metadatos juegan un papel importante en la recuperación de la información, lo que se debe en gran medida a que estos son considerados “datos altamente estructurados”; significa que hay una rigurosidad y estructuración para cada componente del metadato, el cual sigue un patrón bajo parámetros específicos de segmento o conjunto, con lo cual se logra homogeneidad en los datos almacenados para enlazar al recurso. Estos parámetros pueden ser los siguientes: semántica, núcleo de metadatos, operadores booleanos y truncadores. A continuación, se describen cada uno de ellos.

1.4.1. Semántica

Según la Real Academia de la Lengua Española, la semántica son los “aspectos del significado, sentido o interpretación de signos lingüísticos como símbolos, palabras, expresiones o representaciones formales”; es la forma en la que cada una de las palabras o textos que se registra en los metadatos cumple su función de representar lo que se quiere mostrar. Es importante tener en cuenta los siguientes aspectos de la semántica en el proceso de creación de metadatos, para que la búsqueda de información conduzca al usuario hacia el documento u objeto de interés:

- ✓ Sinonimia: se refiere a dos palabras de similar significado, pero de escritura distinta, p. ej. barco y buque.
- ✓ Polisemia: son palabras con diferentes significados y similar escritura, p. ej. sierra, el cual puede significar un accidente geográfico o una herramienta de sesgado.
- ✓ Homonimia: son palabras de tono de lectura, pero con escritura distinta, p. ej. sabia y savia.

1.4.2. Núcleo de metadatos

Es el formato basado generalmente en un estándar, que define la estructura, los campos y la forma de ingresar los textos que dan el significado al metadato. A manera de ejemplo, a continuación, se describe el núcleo de Dublin Core (DC) considerado el estándar más usado para la representación y recuperación de información electrónica, debido a su adaptabilidad a los cambios de internet y su aplicación multidisciplinar apoyada en gran medida por la cooperativa *Online Computer Library Center* (OCLC), y por *World Wide Web Consortium* (W3C).

En un principio, DC fue concebido como un mecanismo de descripción generado por el propio autor del documento entendido como *Document Library Object* (DLO), pero su avance lo ha convertido en un “foco de atención de comunidades involucradas en la descripción formal de recursos o unidades de información (museos, bibliotecas, agencias gubernamentales e incluso organizaciones comerciales)” (Méndez Rodríguez, 2002). Por tal motivo, en la actualidad, es un tópico en constante desarrollo por parte de las comunidades internacionales de las ciencias de la información.

El núcleo de metadatos DC está conformado por quince elementos básicos que pueden ampliarse u omitirse dependiendo de las necesidades del objeto (Tabla 1). A diferencia de la estructura del formato MARC, todos los componentes del DC pueden repetirse las veces que sean necesario, debido a los diversos estados y estructuras que posee la información electrónica.

Tabla 1. Elementos del estándar internacional Dublin Core².

Elemento DC	Equivalencia	Descripción y uso
<i>DC. Title</i>	Título	Nombre dado a un documento usualmente por el autor.
<i>DC. Subject</i>	Materia	Materias de recurso. Se expresan los descriptores o frases que describen el contenido del DLO. Se fomenta el uso de vocabularios controlados y de sistemas de clasificación o <i>schemes</i> .
<i>DC. Description</i>	Descripción	Descripción del recurso; un resumen, en el caso de un documento de texto, o una descripción del contenido en el caso de un documento visual.
<i>DC Source</i>	Fuente	Secuencia de caracteres utilizada para identificar unívocamente un trabajo a partir del cual proviene el recurso actual.
<i>DC. Language</i>	Idioma	Idioma(s) del contenido intelectual del DLO. El contenido de este campo debe coincidir con la norma ISO 639 (RFC 1766).
<i>DC. Relation</i>	Relación con otros documentos	Identificador de un recurso y de la relación que tiene con el objeto de información que se está describiendo. Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos. Por ejemplo: <i>IsVersionOf</i> (Versión de un trabajo), <i>IsBasedOn</i> (basado en traducción, arreglos, etc.), <i>IsPartOf</i> (Es parte de un componente mayor), <i>IsFormatOf</i> (Transformación de un conjunto de datos).
<i>DC. Coverage</i>	Cobertura	Característica de cobertura espacial y/o temporal del contenido del recurso. Se refiere a la región física [...] generalmente son datos extraídos de una lista controlada o <i>scheme</i> . La cobertura temporal se refiere a la temporalidad del objeto.
<i>DC. Creator</i>	Creador, Autor	Persona física o jurídica responsable de la creación del contenido intelectual del recurso. Por ejemplo, los autores, en el caso de documentos escritos, y artistas, fotógrafos e ilustradores, en el caso de recursos visuales.
<i>DC. Publisher</i>	Editor	Entidad responsable de hacer que el recurso se encuentre disponible en su formato actual. (p. ej., la empresa editora, una universidad u otro tipo de organización).
<i>DC. Contributor</i>	Colaborador	Persona u organización que haya tenido una contribución intelectual significativa en la creación del documento. (p. ej., editor, ilustrador, traductor, etc.).
<i>DC. Rights</i>	Derechos	Referencia sobre derechos de autor, como por ejemplo, URL o enlace a un bien, a un servicio de gestión de derechos o a un servicio que dará información sobre los términos y condiciones de acceso al recurso electrónico o DLO.
<i>DC. Date</i>	Fecha	Fecha en la que el recurso se puso a disposición de los usuarios en su forma (electrónica) actual. Esta fecha no debe confundirse con el elemento <i>DC. Coverage</i> que se asocia a un recurso solo en la medida en que el contenido intelectual del mismo está relacionado con esa fecha. Para determinar la fecha de publicación o de disposición al público se utiliza el formato W3CDTF (<i>Consorcio Web, Data and Time Format</i>).
<i>DC. Type</i>	Tipo de recurso	Categoría del DLO en cuanto al tipo de información que representa (p. ej., imagen, conjunto de datos, texto, software, etc.).
<i>DC. Format</i>	Formato	Formato de datos del DLO, usado para identificar el software y posiblemente, también el hardware que se necesita para mostrarlo.
<i>DC. Identifier</i>	Identificador	Secuencia de caracteres usados para identificar unívocamente un recurso. Para los recursos en línea, estos identificadores serán URI (URL/URN) u otras tentativas individuales (DOI, WEBDAVC, OpenURL, etc.) válidas para identificar de manera uniforme un DLO. Para otros recursos pueden usarse formatos de identificadores como ISBN, ISSN, etc.

² Traducido y adaptado de: DMI. *Dublin core Metadata Initiative* (Revisado el 16 de octubre de 2019). <https://www.dublincore.org/>.

1.4.3. Operadores booleanos

La recuperación de información en un modelo de metadatos, reconfigura cada término o palabra para que se generen relaciones con otros registros de metadatos y de esta forma recuperar varios elementos simultáneamente. Todo esto se puede lograr mediante el uso de varios modelos, entre ellos:

- El modelo básico de sistema de búsqueda de información, que usa una jerarquización temática en la cual se genera una reacción basada en árboles, con un eje temático central que despliega diferentes ramas temáticas cada vez más específicas dependiendo de las necesidades del sistema.
- El modelo complejo basado en la búsqueda por palabras o descripciones más específicas en registros documentales de una base de datos, logradas por los operadores booleanos de proximidad y de tiempo.

Sobre este último, la matemática booleana desarrollada en el siglo XIX por George Boole definió condicionales de verdadero y falso en diferentes cálculos matemáticos; con el surgimiento de las bases de datos estas fueron aplicadas a las búsquedas de información³, usando la semántica en tres conectores principales (Figura 2) que se describen a continuación:

- **AND** también utilizado como **Y**: define que, si hay presencia de dos o más palabras buscadas en el registro de metadatos, lo recupere, en caso contrario lo omita. Es muy útil para acotar una consulta. Ejemplo: océanos AND Colombia.
- **OR** también utilizado como **O**: este conector define que, si hay presencia de una de las dos o más palabras registradas en el metadato, lo recupere, en caso que no exista ninguna lo omita. Es útil para recuperar grandes conjuntos de metadatos para refinar con otros conectores. Ejemplo: puertos OR capitanías.
- **NOT** también utilizado como **NO**: este conector es excluyente, define que si hay presencia del primer término y no hay presencia del segundo lo recupere, en caso que sí lo omita. Este es inverso al operador **AND**. Ejemplo: oceanografía NOT 2018.

³ En algunos sistemas los operadores booleanos han sido incluidos gráficamente como botones de filtrado para búsquedas específicas o avanzadas de información.

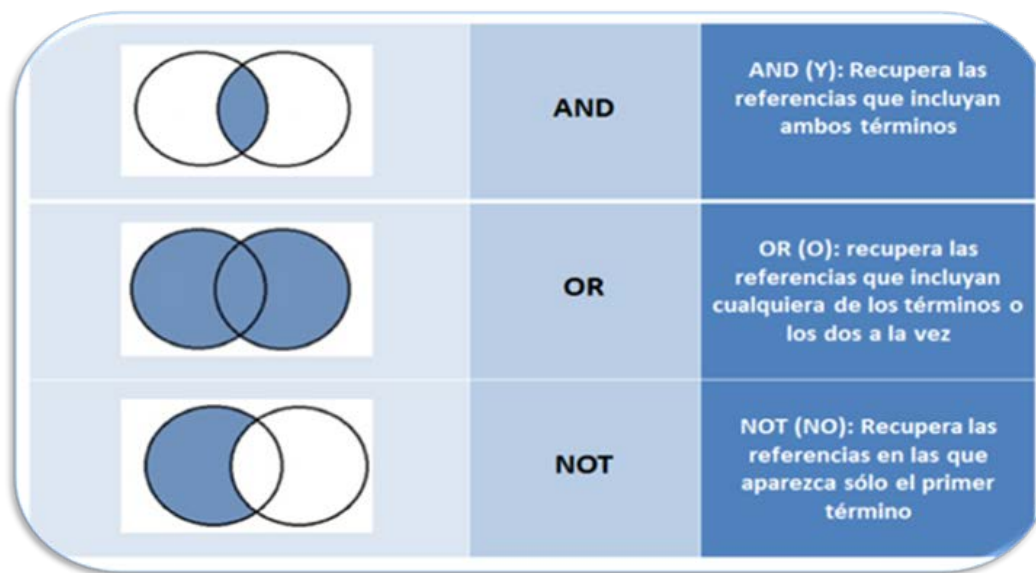


Figura 2. Ilustración del uso de los operadores booleanos⁴.

1.4.4. Truncadores

Al igual que los operadores booleanos, estos son comodines de caracteres especiales diseñados en los años 80 del siglo XX con el nacimiento de los sistemas de búsqueda electrónicos y las bases de datos, para direccionar la búsqueda por un patrón específico; los más comunes son:

- **Comillas “ ”**: al utilizar estos caracteres el sistema solo recuperará registros que cuyos textos en el metadato coincidan con la frase de búsqueda, en caso contrario los omitirá aún si las palabras se encuentren de forma dispersa en el metadato. Ejemplo: “marina mercante”.
- **Asterisco ***: Busca todo registro que contengan palabras en el contenido de los metadatos de la misma familia léxica con diferencia en uno o varios caracteres. Ejemplo: Para la búsqueda ocean*, el sistema retornará registros con palabras tales como océano, océanos u oceanografía.
- **Paréntesis ()**: Este conector crea conjuntos con los cuales segmenta una búsqueda usando la propiedad asociativa y la teoría de conjuntos. Dentro de cada conjunto es posible utilizar tanto los operadores booleanos como los truncadores. Ejemplo: (“marina mercante” AND Colombia) NOT (leyes OR ocean*).

⁴ Fuente: El autor.

1.5. Reflexiones finales

Los metadatos están presentes en la vida cotidiana, en algo tan simple como un documento de identidad con cuyos datos y mediante los sistemas de información, se pueden generar puntos en común de millones de estos en diferentes épocas, temáticas y finalidades. Es así que los datos e información que se producen pueden ser más útiles a la comunidad, si existe un riguroso y normalizado registro de metadatos.

Los metadatos contribuyen a resolver la necesidad de acceso a datos e información, sea por parte de un individuo o de una comunidad; con sus elementos es posible optimizar los procesos de búsqueda y de indexación que permiten llegar a un documento por cualquier concepto, como una frase exacta o palabras sueltas.

2. EXPERIENCIAS ►

2.1. Arqueología y recuperación de datos recopilados durante más de 40 años a bordo de cruceros oceanográficos colombianos

2.1.1. Creación del Centro Colombiano de Datos Oceanográficos

Las actividades de investigaciones marinas en forma periódica en Colombia se empezaron a ejecutar en 1969 en el marco de lo establecido en el Decreto 763 de ese mismo año, por medio del cual se creó la Comisión Colombiana de Oceanografía (actualmente Comisión Colombiana del Océano, CCO). Después de cuatro años de actividad de campo en el mar, ya era considerable el volumen de datos oceanográficos recopilados, así como el número de usuarios que requerían acceder a ellos para diferentes fines (Dimar, 1973).

Con lo anterior, se evidenció la necesidad de aplicar las recomendaciones de la Comisión Oceanográfica Intergubernamental (COI) de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco, por sus siglas en inglés) sobre administración de datos y satisfacer la necesidad de acceso a estos (Dimar, 1992). Es así que, en 1973 la Sección de Estudios Especiales de la División de Oceanografía de la entonces Dirección General Marítima y Portuaria (Dimar), formuló el proyecto “Centro Colombiano de Datos Oceanográficos - Cecoldo” (Dimar, 1973).

El inicio de la ejecución del proyecto en mención fue planeado para el 1 de septiembre de 1973 con una duración de dos años (cada año una fase), sin embargo, el Cecoldo contó con la capacidad de prestar algunos de sus servicios antes de completar la segunda fase del proyecto, tal como se evidencia en el informe del Crucero Oceanográfico “Océano III” desarrollado entre abril y mayo de 1974 en el Caribe colombiano, en el cual se nombra formalmente al Cecoldo como encargado del procesamiento de datos “por medio de un programa de computador en Fortran IV [...] y un programa para estaciones oceanográficas del *National Oceanographic Data Center* (NODC) de los Estados Unidos” (Dimar, 1975), así como en los informes sucesivos de cruceros oceanográficos hasta 1984 (Figura 3).

A finales de 1978 el Cecoldo adquirió su propia capacidad de cómputo al recibir de la Organización de Estados Americanos (OEA) en el marco del Programa Multinacional de Ciencias del Mar, un minicomputador WANG, y gradualmente fue ganando independencia del Departamento Administrativo Nacional de Estadística (DANE), en cuyos equipos se procesaron los primeros años de datos obtenidos en cruceros oceanográficos colombianos (Dimar, 1992).

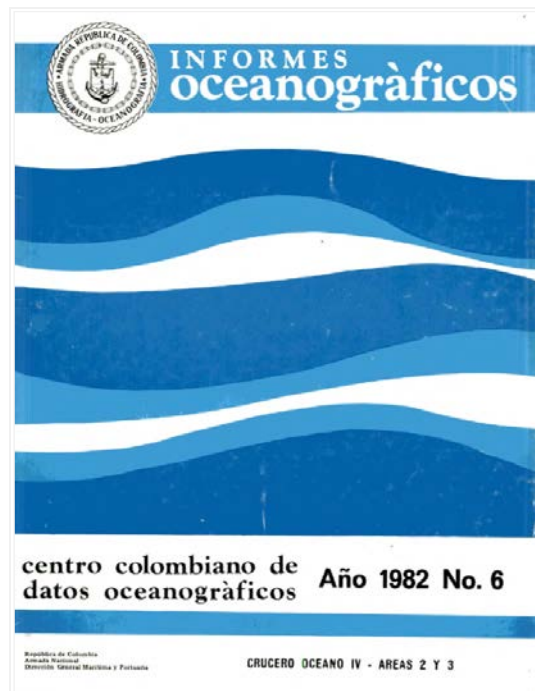


Figura 3. Portada del Informe de Crucero Oceanográfico “Océano IV (Áreas 2 y 3) – 1974” publicado en 1982⁵.

2.1.2. La primera experiencia en arqueología y recuperación de datos

Debido a la falta de personal y obsolescencia de equipos de cómputo utilizados, en 1984 se registró la última actividad formal del Cecoldo con el procesamiento de los datos del Crucero ERFEN⁶ IV desarrollado entre el 29 de noviembre y el 12 de diciembre de 1978 en el Pacífico colombiano. Siete años más tarde, por orden del señor Director General Marítimo se asignó el Cecoldo a la dependencia orgánica de la Secretaría General de Dimar (y no de la Sección de Sistemas como se había concebido inicialmente), y se reactivó el 31 de enero de 1992, dando inicio al procesamiento y control de calidad de los datos obtenidos en 23 cruceros oceanográficos, de 38 desarrollados hasta la fecha en Colombia (Dimar, 1992).

La metodología que se aplicó en esta primera recuperación de datos oceanográficos incluyó reunir y documentar información sobre los siguientes aspectos:

- **Resumen de la operación:** generalidades del crucero tales como nombre, buque oceanográfico utilizado, periodo de tiempo y descripción del área de operación.
- **Muestreo:** descripción general de las variables y métodos de medición o de muestreo.

⁵ Fuente: Dimar

⁶ Estudio Regional del Fenómeno El Niño/La Niña.

- **Descripción de formato:** descripción detallada de las variables, unidades de medida, instrumentos y/o tablas de códigos o convenciones utilizados.
- **Notas:** límites de detección, explicación del uso de caracteres alfabéticos o numéricos, problemas presentados durante el muestreo.
- **Red de estaciones:** muestra gráfica de la malla de estaciones de muestreo.
- **Gráficos y datos:** muestra gráfica de la distribución de parámetros oceanográficos observados durante el crucero, así como el valor de los datos primarios.

2.1.3. Implementación de “OceanData”

En 1996 el Centro de Investigaciones Oceanográficas e Hidrográficas (CIOH) asumió el Centro Nacional de Datos Oceanográficos (NODC, por sus siglas en inglés) ante la COI, y “consiente de la necesidad de manejo adecuado que debe darse a la información producida” (CIOH, 2002) implementó las fases I (1996) y II (1997) del sistema de información oceanográfico e hidrográfico denominado “OceanData” (Figura 4). Posteriormente en 2001, desarrolló la fase III en el marco del programa para el Intercambio Internacional de Datos e Información Oceanográfica (IODE) de la COI, “en respuesta a las necesidades de manejo y divulgación de información tanto a nivel interno como externo” (CIOH, 2002). Cabe anotar que no se obtuvo financiación para ejecutar la fase IV del proyecto “OceanData” y con los años este sistema fue relegado.



Figura 4. Aspecto de uno de los módulos de “OceanData” del CIOH⁷.

La metodología aplicada durante la recuperación de datos efectuada en el marco de la implementación de “OceanData”, estuvo enfocada al control de calidad de los datos de salinidad y temperatura del agua obtenidos en los cruceros oceanográficos “Caribe” y “Océano” desarrollados hasta 1993 (CIOH, 2002). Las pruebas aplicadas fueron las siguientes:

⁷ Fuente: CIOH (2002).

- Análisis visual de los datos para detectar y eliminar, entre otros, errores en la utilización del instrumento, como los de tiempo de bajada o subida del mismo.
- Suavización de los datos aplicando la rutina WFILTER del software SEASOFT del instrumento CTD Sea-bird.
- Selección de muestreos espacialmente, aplicando la rutina BINAVG para determinar los datos a profundidades metro a metro.
- Filtrado de datos a profundidades estándar (NOAA, 2015)⁸ y exportación de los archivos al formato XLS.
- Cargue de información descriptiva sobre los cruceros oceanográficos en OceanData.

2.1.4. Las iniciativas “Cenimarc” y “Retrocean”

En 2004, el entonces Centro Control Contaminación del Pacífico (CCCP⁹) fue designado como NODC y dio inicio al proyecto “Central de Información Marina Colombiana, Cenimarc”, con el objetivo de “proporcionar una solución sistemática a los problemas representados por la redundancia de la información y el tratamiento no estructurado de los datos generados por los centros de investigación de la DIMAR” (Ortiz y Rico, 2006).

En el marco de dicha iniciativa, el CCCP adelantó actividades de arqueología y recuperación de datos de los cruceros oceanográficos desarrollados en el Pacífico colombiano, destacándose las siguientes actividades (CCCP, 2004):

- Ubicación de fuentes de datos en las instalaciones del CCCP.
- Diseño de plantillas en formato XLS para importar los datos.
- Desarrollo de una aplicación en formato XML para cargar los datos en Cenimarc desde las plantillas XLS.
- Desarrollo de una función para exportar los resultados de una consulta al formato *Ocean Data View* (ODV).

⁸ Son profundidades expresadas en metros correspondientes a los niveles definidos para el intercambio internacional de mediciones oceanográficas en la columna de agua. Estas son: 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2500, 3000, 3500, 4000, 4500, 5000 y 5500 m (NOAA, 2015).

⁹ Mediante el artículo 6 del Decreto 5057 de 2009, el nombre del CCCP cambió a “Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico”.

- Implementación del sistema de administración de metadatos *Marine Environmental Data Information Referral Catalogue* (MEDI).

Por su parte, en 2005 el CIOH desarrolló la primera fase del proyecto “Retrocean” que planeó la “recuperación y control de calidad de datos oceanográficos”, y específicamente “información oceanográfica de los diferentes cruceros desarrollados por el CIOH en más de 30 años de investigación en el mar Caribe, para que esta fuera consignada en la base de datos de Cenimarc con los respectivos metadatos” (CIOH, 2005). En la Figura 5 se aprecia el resumen de las actividades de arqueología y recuperación desarrolladas en Retrocean.



Figura 5. Metodología empleada en la fase I del proyecto “Retrocean”¹⁰.

Tanto la segunda fase del proyecto Retrocean como las fases sucesivas de Cenimarc no recibieron financiación, por lo que no fue posible llevar a término las actividades planeadas.

¹⁰ Fuente: Adaptado de CIOH (2005).

2.1.5. El Atlas de los Datos Oceanográficos de Colombia

En 2013 se firmó un convenio específico de cooperación entre Dimar, Ecopetrol y la Asociación Colombiana para el Avance de la Ciencia (ACAC), que tuvo por objeto “aunar esfuerzos para realizar el Atlas Oceanográfico de Colombia mediante la recopilación, organización, depuración, sistematización y análisis de los datos oceanográficos colectados por Dimar, para consulta a través de medios análogos por parte de la comunidad marítima nacional” (Convenio 5211792, 2013).

Como resultado, se publicaron en 2015 las climatologías del periodo 1992-2013 para la temperatura, salinidad y densidad en niveles estándar de profundidades seleccionadas, a partir de datos obtenidos a bordo de los buques oceanográficos de la Armada de la República de Colombia (ARC) y Dimar en la cuenca Caribe y cuenca Pacífica de Colombia, y los disponibles en el NODC y el Centro Mundial de Datos (WDC) de la Administración Nacional Oceánica y Atmosférica (NOAA, por sus siglas en inglés) de los Estados Unidos (Andrade *et al.* 2015).

En la Figura 6 se aprecia el flujo de trabajo seguido en el procesamiento de los datos para producir el “Atlas de los Datos Oceanográficos de Colombia: 1922-2013 Temperatura, Salinidad, Densidad, Velocidad Geostrofica”, el cual incluyó actividades de arqueología y recuperación tales como el inventario y digitalización de datos y la implementación de una base de datos para una aplicación software de escritorio. La descripción metodológica y la climatología puede ser consultada en: <http://cecoldodigital.dimar.mil.co/49/>.

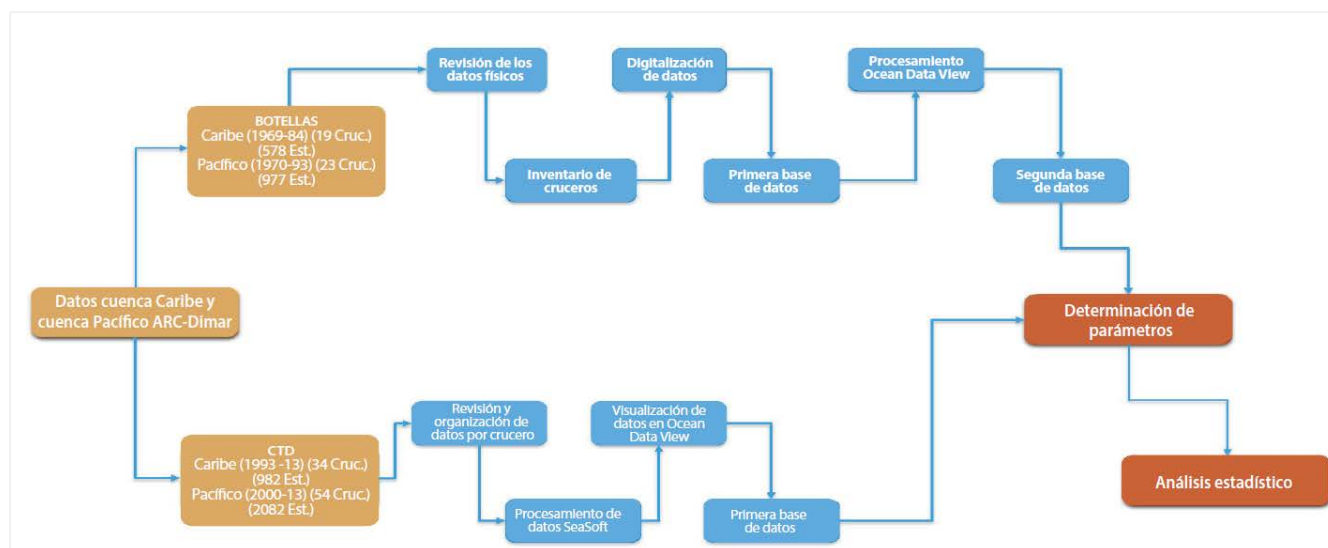


Figura 6. Diagrama de flujo seguido en el procesamiento de datos oceanográficos¹¹.

¹¹ Fuente: Andrade *et al.*, (2015).

2.1.6. Posicionamiento estratégico del Cecoldo

En 2014 se retomaron los esfuerzos de arqueología y recuperación de datos primarios realizados por el “Atlas de los Datos Oceanográficos de Colombia”, y como una iniciativa de la Subdirección de Desarrollo Marítimo de Dimar se formuló el proyecto de inversión “Posicionamiento estratégico del Cecoldo” ejecutado formalmente entre 2015 y 2017. Este proyecto tuvo por objeto “conocer la cantidad y la calidad de los datos oceanográficos e información marina producidos por Dimar a partir de actividades de investigación y monitoreo del Pacífico y Caribe colombianos” (Dimar, 2018).

Dentro de los objetivos específicos del proyecto se planeó realizar la arqueología, recuperación y estandarización de datos oceanográficos y de meteorología marina producidos por Dimar entre el periodo 1969 y 2017. La metodología aplicada para lograr este objetivo correspondió a la recomendada por Hernández *et al.*, (2007) (Figura 7), complementando la “FASE III. Conservación” con la “Guía para transferencia y disposición final documental” (Dimar, 2016) del proceso de gestión documental y de contenido de Dimar, la cual incluye los lineamientos emitidos por el AGN en la materia.

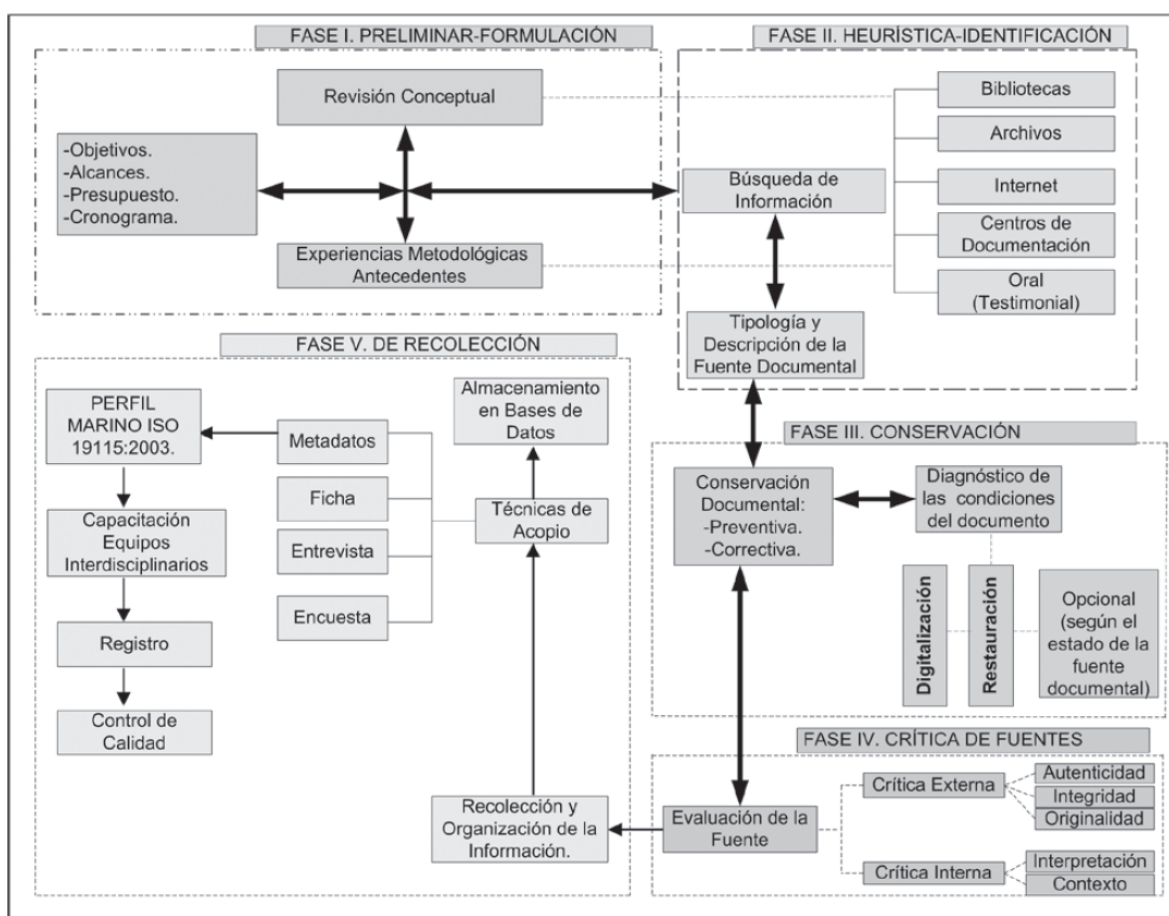


Figura 7. Fases de la metodología archivística para la recuperación de información histórico-oceanográfica propuesta por Hernández *et al.*, (2007).

La búsqueda y tratamiento de documentos históricos de datos e información (en soporte papel y electrónico) se llevó a cabo en las instalaciones del Centro de Investigaciones Oceanográficas e Hidrográficas del Caribe (CIOH) ubicado en Cartagena D.T. y C. (Figura 8), en el Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico (Cccp) ubicado en Tumaco (Figura 9), y en la Subdirección de Desarrollo Marítimo de Dimar en la ciudad de Bogotá D.C., Colombia.



Figura 8. Imágenes del proceso de la arqueología de datos oceanográficos en el Cioh¹².



Figura 9. Imágenes del proceso de la arqueología de datos oceanográficos en el Cccp¹³.

¹² Fuente: Dimar

¹³ Ídem

Una vez obtenidos e inventariados los datos en formato digital, se procedió con la estandarización aplicando la “Guía normalización de conjuntos de datos oceanográficos” del Cecoldo (Dimar, 2017), en la cual se aplican estándares y mejores prácticas recomendadas por el programa internacional IODE de Unesco-COI, entre los cuales se destacan la aplicación de normas ISO, formatos, vocabularios y la documentación y publicación de metadatos.

Cabe anotar que el control de calidad de datos efectuado en este proyecto, se enfocó a garantizar que los conjuntos de datos cumplieran con los estándares acordados, y no a la evaluación o validación de cada uno de los parámetros que los componen (es decir, la temperatura, salinidad, oxígeno, etc.), por tanto, no se asignaron banderas de calidad que califican cada dato obtenido como malo, bueno o cuestionable. Este último es objeto de iniciativas que adelanta el Cecoldo en la actualidad.

Como se aprecia en la Tabla 2, se recuperaron datos oceanográficos (físicos y químicos) y de meteorología marina de 118 cruceros oceanográficos, con la cobertura geográfica observada en la Figura 10 cuyos muestreos estuvieron a cargo de la Armada de la República de Colombia o de los Centros de investigaciones de Dimar. La mayor cantidad de cruceros han sido desarrollados en la cuenca del Pacífico colombiano y más recientemente en áreas insulares como Isla de Malpelo e Isla Gorgona, en el marco del Estudio Regional del Fenómeno El Niño/La Niña (Erfen), para un total de 63 en un periodo de 46 años, seguido de los Cruceros Oceanográficos Caribe con un acumulado de 18 en un lapso de 21 años.

Tabla 2. Cantidad de cruceros cuyos datos oceanográficos fueron recuperados en el marco del proyecto “Posicionamiento estratégico del Cecoldo”¹⁴.

Expedición	Área geográfica	Cantidad	Años o periodo
Crucero Oceanográfico Caribe	Caribe	18	1992 – 2013
Crucero Oceanográfico CICAR ¹⁵	Pacífico	3	1972 – 1973
Crucero Oceanográfico Golfo de Urabá	Caribe	3	2005 y 2010
Crucero Oceanográfico Guajira	Caribe	2	1990 y 2010
Crucero Oceanográfico Islas del Rosario	Caribe	3	1977 – 1978
Crucero Oceanográfico Océano	Caribe	9	1969 – 1984
Crucero Oceanográfico Pacífico - Erfen	Pacífico	63	1970 – 2016
Crucero Oceanográfico Petrobras	Caribe	4	2007 – 2008
Crucero Oceanográfico San Andrés y Providencia	Caribe	1	2010
Crucero Oceanográfico SPOA ¹⁶	Caribe	5	2007 – 2008
	Pacífico		
Crucero Oceanográfico y Pesquero	Caribe	3	1990 y 1996
	Pacífico		
Expedición Científica Seaflower	Caribe	2	2014 y 2016
Expedición Científica de Colombia a la Antártica	Antártica	2	Verano austral 2014-2015 Verano austral 2016-2017

¹⁴ Fuente: Dimar-Cecoldo.

¹⁵ Programa ‘Investigaciones Cooperativas del Caribe y Regiones Adyacentes’.

¹⁶ Proyecto ‘Sistema de Pronósticos Oceánicos y Atmosféricos’.

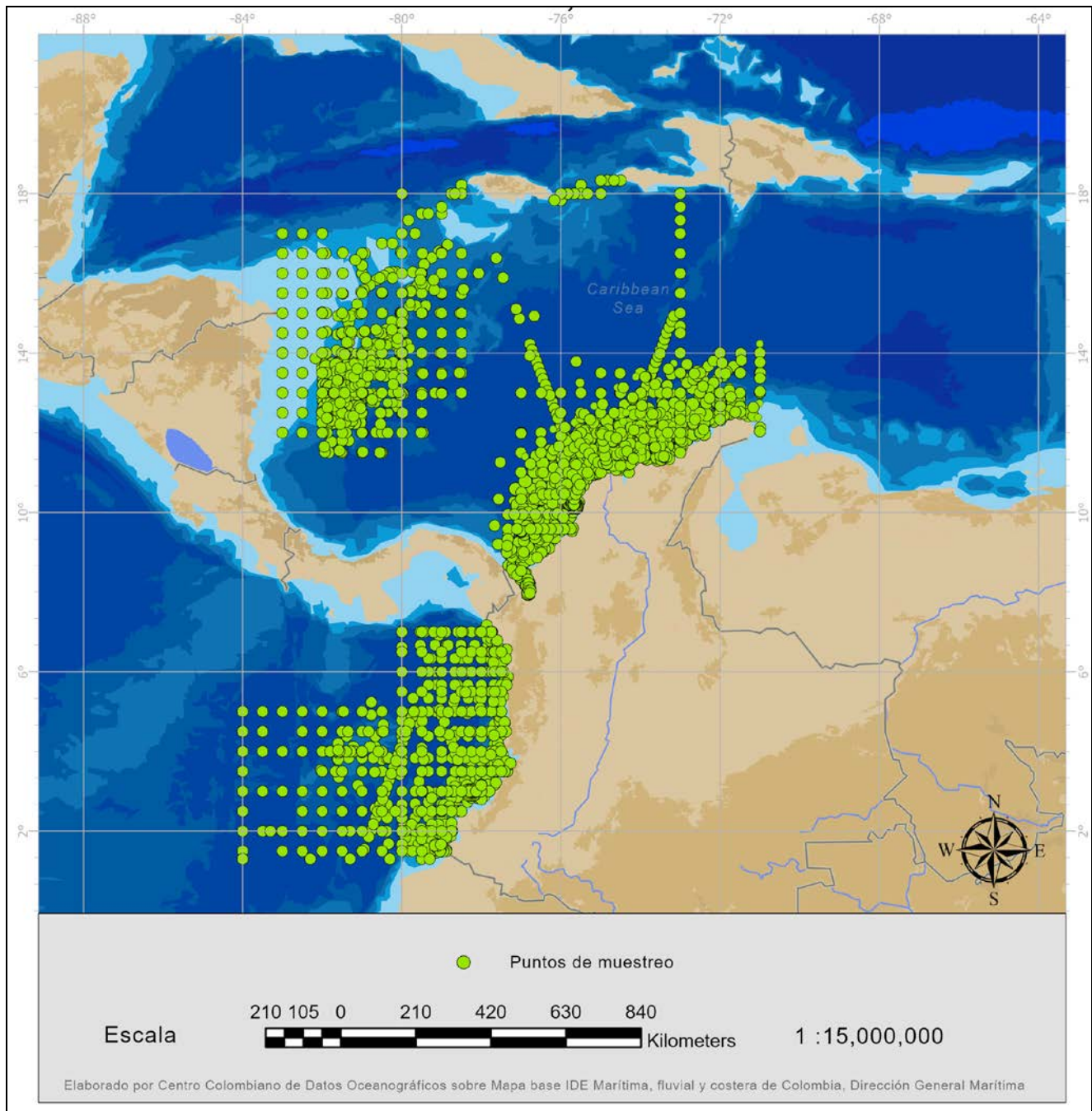


Figura 10. Cobertura geográfica de los conjuntos de datos oceanográficos y de meteorología marina 1969-2017 recuperados en el proyecto “Posicionamiento estratégico del Cecoldo”¹⁷.

¹⁷ Fuente: Dimar-Cecoldo

2.1.7. Conclusiones

El grado de conservación de la documentación en soporte papel, tales como informes de crucero, informes de proyectos y carpetas con hojas de datos, permitió contar con gran parte de la fuente primaria de los datos históricos de cruceros oceanográficos. Lo anterior se debe en gran medida a los procesos de gestión documental que las instituciones han venido implementando en los últimos años con la dirección del AGN, y que han permitido que el archivo histórico se conserve bajo condiciones específicas. Solo en pocos casos no se recuperaron datos biológicos y geoquímicos ya que el mencionado componente estuvo a cargo de otras instituciones o universidades invitadas a los cruceros.

Se observó que, en gran medida, el éxito de este tipo de iniciativas dependió de la continuidad que se le dio al proceso (es decir, la posibilidad de desarrollar la totalidad de los objetivos o fases planeadas), por supuesto, soportado en un apoyo financiero y de disponibilidad de talento humano con dedicación exclusiva a las actividades de arqueología y recuperación.

Se encontró una constante en cada experiencia y consistió en repetir el proceso inmediatamente anterior; en pocas oportunidades se pudo retomar un trabajo desarrollado con antelación, probablemente debido a cambios en las tecnologías o por ausencia del respaldo de los datos ya procesados. Pese a ello, en la actualidad el país cuenta con un repositorio único de datos oceanográficos colombianos, sustentado en lineamientos y protocolos internacionalmente reconocidos en esta disciplina, lo cual es reconocido por los países que componen la región del Pacífico sudeste.

Para una siguiente iniciativa de arqueología y recuperación de datos oceanográficos, se pretende ampliar la búsqueda de información en entidades que hayan participado en los cruceros oceanográficos, así como generar vínculos con los productos desarrollados a partir de los mencionados datos. Lo anterior, con el fin de que los usuarios puedan tener una visión global del alcance e impacto que han tenido estas importantes expediciones científicas en el mar.

2.2. Desafíos en la arqueología y recuperación de datos de la red de monitoreo meteomarinero más grande de Colombia

La Dirección General Marítima (Dimar) inició en 2005 la implementación de la Red de Medición de Parámetros Oceanográficos y de Meteorología Marina (RedMpomm), con el fin de mejorar la capacidad de observación y predicción climática a diferentes escalas, así como el monitoreo de las condiciones meteomarineras que puedan afectar la seguridad integral marítima y portuaria del país.

La RedMpomm en la actualidad está compuesta por 35 estaciones de monitoreo meteorológico y/o mareográfico en tiempo real, desplegadas a lo largo del litoral Caribe y Pacífico y áreas insulares colombianas, generando en promedio más de 50.000 datos diarios. Las primeras estaciones fueron instaladas en Turbo, Providencia y Coveñas en el año 2005; en los años posteriores se situaron nuevas plataformas de observación, pero fue en 2013 cuando Dimar realizó la mayor inversión con la instalación de dos estaciones mareográficas y once meteorológicas-mareográficas, con el objetivo de ampliar la cobertura del monitoreo del nivel del mar, como insumo clave para la evaluación del arribo de olas de tsunami a las costas colombianas por parte del Centro de Alerta por Tsunami de Dimar.

Más de una década después de haber iniciado las mediciones, se evidenció que la falta de adopción de estándares y buenas prácticas en todo el ciclo de vida del dato, es decir, planeación, adquisición, procesamiento, archivo y publicación (CTN Diocean, 2018), había llevado a la red a contar con datos duplicados, diversos formatos de almacenamiento, ausencia de datos en las series, y un nivel de calidad que no tenía aceptación por parte de los usuarios internos y externos.

Por lo anterior, y teniendo en cuenta que “los principales objetivos de la gestión de bases de datos consisten en mantener permanentemente la integridad de la base de datos y garantizar que esta contenga todos los datos y metadatos necesarios para responder a las necesidades por cuyo motivo fue establecida, tanto en el presente como en el futuro” (OMM, 2011), Dimar apropió recursos y esfuerzos para realizar la arqueología y recuperación de datos de la RedMpomm. En los siguientes apartados se describen las actividades desarrolladas, las cuales se resumen en las cinco etapas que se aprecian en la Figura 11.

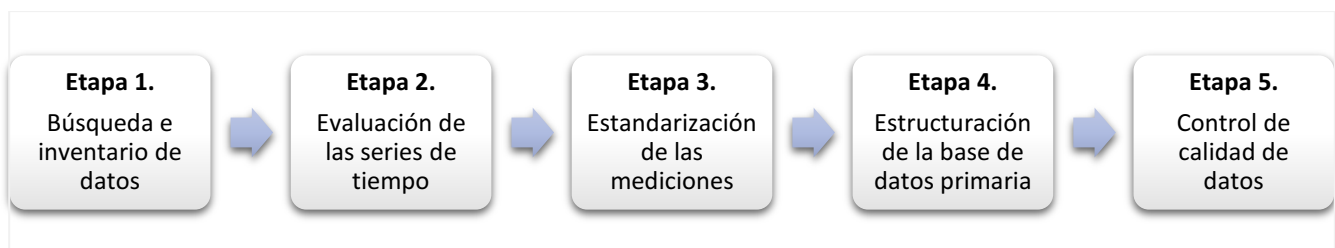


Figura 11. Etapas del proceso de arqueología de datos meteorológicos y del nivel del mar de la RedMpomm¹⁸.

¹⁸ Fuente: Dimar.

2.2.1. Etapa 1. Búsqueda e inventario de datos

En esta etapa se realizó la búsqueda e inventario de los datos encontrados en diferentes equipos de cómputo y servidores, tanto de la sede central de Dimar en Bogotá D.C., como de sus regionales en Tumaco y Cartagena; adicionalmente, se revisaron los dispositivos magnéticos y ópticos (p. ej. discos, CD, DVD) en los cuales se entregaban los informes de mantenimiento preventivo de las estaciones meteorológicas y mareográficas de la RedMpomm.

Los resultados de la búsqueda en mención arrojaron diversas fuentes de datos; estas fueron:

- **Bases de datos nativas.** Este tipo de base de datos corresponde a una serie de directorios estructurados que son gestionados a través del software propio del instrumento o plataforma, en este caso Hydras 3.
- **Raw data¹⁹ de las estaciones.** Hace referencia a los archivos nativos codificados, descargados en sitio cada vez que se realizaba un mantenimiento a las estaciones.
- **Datos exportados.** Corresponde a datos en formatos TXT o XLS, extraídos de la base de datos primaria del sistema para ser utilizados en procesos específicos.

2.2.2. Etapa 2. Evaluación de las series de tiempo

En esta etapa se realizó la verificación de la integridad y disponibilidad de las bases de datos nativas en operación; de igual forma se revisó la continuidad de las mediciones históricas comparando los datos recuperables de las diferentes fuentes identificadas.

Durante este proceso, se encontraron diferentes aspectos que dificultaron la consolidación de las series de tiempo de las variables meteorológicas y mareográficas medidas por la RedMpomm; estas fueron:

- **Datos primarios incompletos.** La *raw data* descargada en sitio se limitaba a periodos de aproximadamente tres meses de datos anteriores a la fecha de mantenimiento; esto debido a la configuración del registrador de datos de las estaciones.
- **Datos faltantes.** Por diversos motivos no se lograron recibir todos los datos de la serie por los medios de telemetría que contaba la Red.

¹⁹ Datos primarios obtenidos de la fuente, en este caso, de las tarjetas de memoria de las plataformas de observación.

- **Bases de datos distribuidas.** Dada la arquitectura distribuida del sistema, se contaba con bases de datos locales para cada grupo de usuarios, las cuales al no estar sincronizadas aumentaron la incertidumbre de las series de datos. Como resultado, no se contaba con una base de datos primaria única y consolidada.
- **Gobierno de datos²⁰.** Al no tener un gobierno de datos acorde a las necesidades de la Red, la operación y conocimiento del dato no estaba soportado en procedimientos, guías o manuales que evidenciaran las transformaciones y usos de estos.

Una vez consolidadas las series de datos, se procedió con la evaluación de las mismas encontrando los siguientes errores:

- **Datos fuera del rango de tiempo.** Se identificaron incoherencias en las fechas de inicio y fin registradas en algunas mediciones ya que no coincidían con los periodos de tiempo de captura basados en la puesta en operación de los sensores; por lo anterior, se encontraron datos de años anteriores a la instalación de las estaciones y posteriores a la fecha límite de verificación (p. ej. datos de fechas anteriores a 2005 y datos de 2115) (Figura 12).

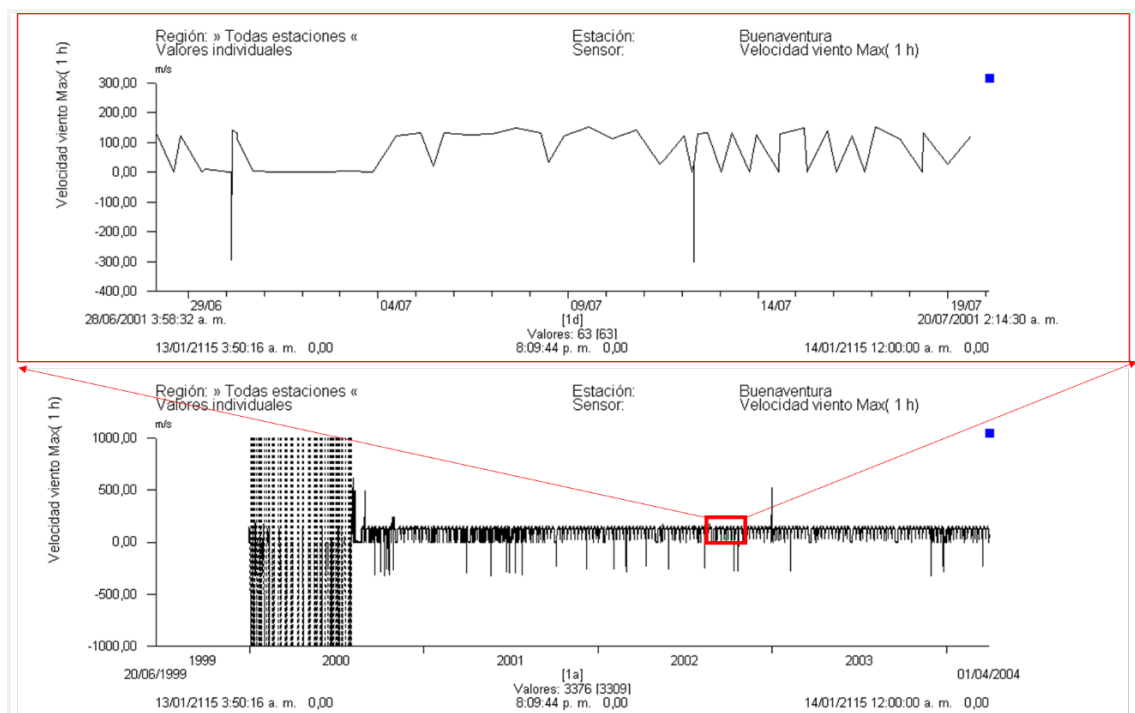


Figura 12. Datos fuera de rango de tiempo en una serie de velocidad del viento para estación de Buenaventura²¹.

²⁰ El Gobierno de datos “puede ser entendido como garantía para los procesos de transformación ‘Eficiencia operativa’, ‘Conocimiento del dato’ y ‘Uso de información’” (MinTIC, 2014).

²¹ Fuente: Dimar-RedMpom.

- **Datos editados.** Ante la ausencia de datos en algunos periodos de tiempo, ya sea por daños en la estación o por novedades en la transmisión de los datos, se intentó estimar los valores de los datos faltantes, pero no se encontró evidencia de la metodología aplicada para generarlos (Figura 13).

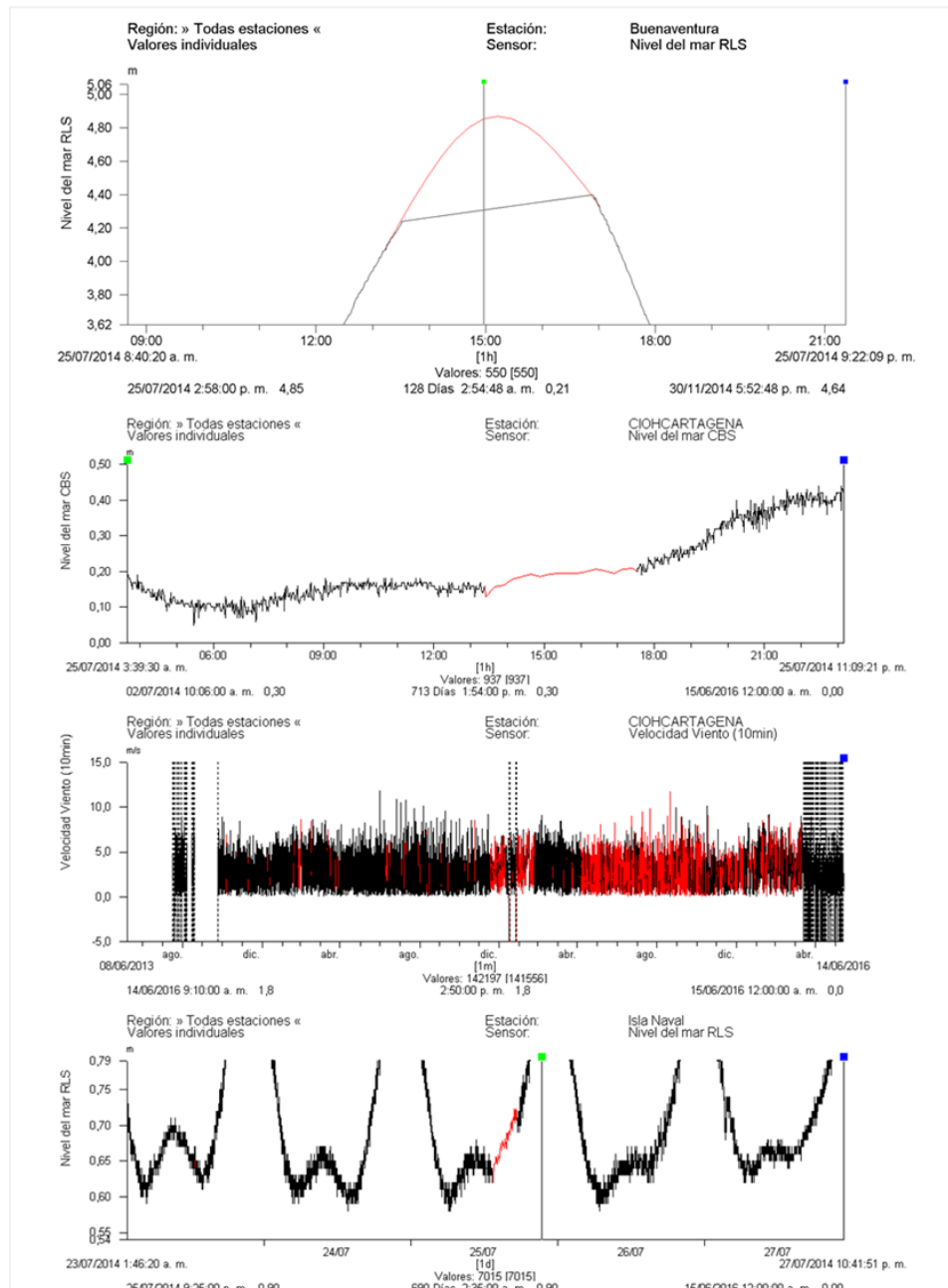


Figura 13. En rojo, datos estimados en series de estaciones ubicadas en Buenaventura y Cartagena²².

²² Fuente: Dimar-RedMppomm.

- **Datos imposibles:** Se identificaron periodos de tiempo en los cuales se registraron datos físicamente imposibles, teniendo en cuenta que estos se encuentran fuera del rango de medición del sensor (Figura 14).

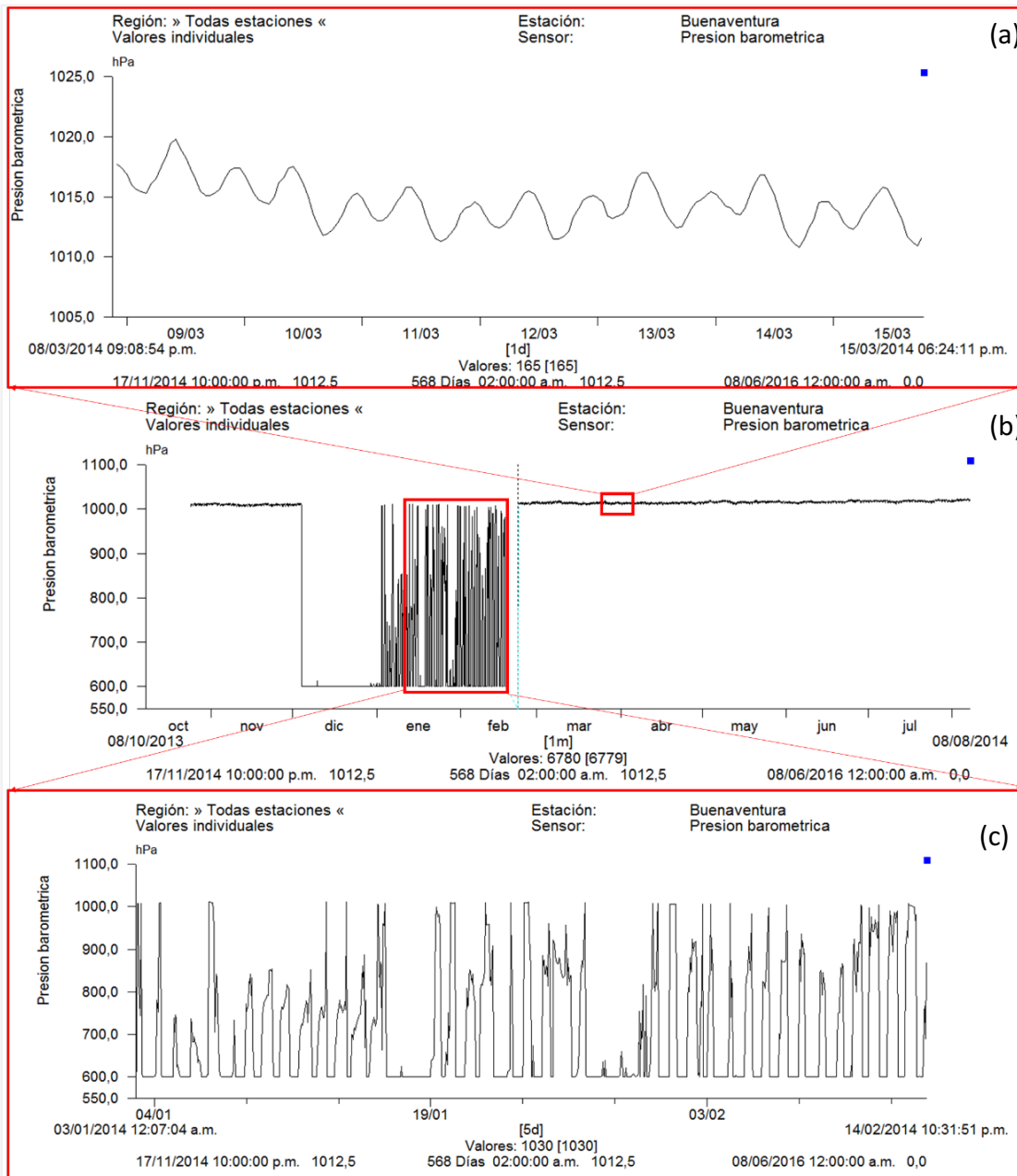


Figura 14. Presión barométrica de estación Buenaventura: (a) detalle del comportamiento o variabilidad esperados; (b) periodo de tiempo de análisis; y (c) detalle de datos imposibles²³.

²³ Fuente: Dimar-RedMppomm.

2.2.3. Etapa 3. Estandarización de las mediciones

Una vez culminada la recuperación y revisión de los datos se procedió con la estandarización de las mediciones, así:

- Se fijaron fechas límites de inicio y finalización de la toma de datos de cada estación, comparando los datos encontrados con la información histórica disponible en las hojas de vida de las plataformas e información contractual.
- Se estandarizaron los códigos de los sensores, se establecieron las agregaciones temporales (es decir, cada cuánto se realizan las mediciones o frecuencia de la medición) y se efectuó la configuración de sensores con la información de la Tabla 3.

Tabla 3. Agregación temporal de los sensores de datos de la RedMpomm²⁴.

Parámetro	Agregación Temporal
Dirección del viento	10 min
Velocidad del aire	10 min
Temperatura del aire	60 min
Humedad Relativa	60 min
Presión barométrica	60 min
Radiación Global	60 min
Precipitación	10 min
Nivel del mar	01 min
Temperatura del mar	10 min

2.2.4. Etapa 4. Estructuración de la base de datos primaria

Con la información recolectada y las decisiones tomadas en etapas anteriores, se procedió a restablecer la base de datos primaria. Para ello, se creó un espacio de trabajo en el software Hydras 3 en el cual se importó la *raw data* encontrada, sirviendo esta como base para la consolidación de los datos.

A continuación, se procedió con la importación de los datos disponibles en el sistema (los cuales previamente se exportaron bajo las configuraciones adoptadas en la etapa 3); esta actividad se acompañó de una inspección visual para identificar vacíos que se pudieran llenar con los datos encontrados en otras fuentes, pero omitiendo datos calculados, procesados o con evidencia de modificaciones o transformaciones.

²⁴ Fuente: Dimar-RedMpomm.

Como resultado de esta etapa, se logró consolidar la base de datos primaria con cerca de 143 millones de datos meteorológicos y de nivel del mar, los cuales incluyen series de tiempo discontinuas y cuyos porcentajes de recuperación se aprecian en la Tabla 4 de acuerdo al tipo de plataforma y lugar de ubicación.

Tabla 4. Datos meteorológicos y de nivel del mar de la RedMpomm recuperados a 31 de diciembre de 2018²⁵.

Tipo de plataforma	Ubicación	% Total de datos recuperados	Serie de tiempo anual
Estación meteorológica	Coveñas	69,75%	13
Estación meteorológica	Providencia	81,36%	13
Estación meteorológica	Turbo	68,62%	13
Estación mareográfica	Buenaventura	64,80%	10
Estación meteorológica	Buenaventura	82,42%	10
Estación meteorológica y mareográfica	Malpelo	79,26%	9
Estación meteorológica	Gorgona	64,20%	8
Estación meteorológica y mareográfica	Tumaco	89,09%	8
Estación meteorológica	Barranquilla	84,83%	7
Estación mareográfica	Puerto Velero	84,04%	6
Estación mareográfica	Cartagena	80,53%	6
Estación mareográfica	Coveñas	91,43%	6
Estación mareográfica	Providencia	97,49%	6
Estación mareográfica	Puerto Bolívar	59,92%	6
Estación mareográfica	Puerto Estrella	74,50%	6
Estación meteorológica	Puerto Bolívar	67,70%	6
Estación meteorológica	Puerto Estrella	78,37%	6
Estación meteorológica	Puerto Velero	88,91%	6
Estación meteorológica	Cartagena	90,98%	6
Estación meteorológica y mareográfica	Bahía Solano	84,45%	6
Estación meteorológica y mareográfica	Ballenas	83,20%	6
Estación meteorológica y mareográfica	Isla Naval	94,55%	6
Estación meteorológica y mareográfica	Juanchaco	84,16%	6
Estación meteorológica y mareográfica	San Andres	93,82%	6
Estación meteorológica y mareográfica	Sapzuro	58,48%	6
Estación meteorológica y mareográfica	Serranilla	29,09%	6
Estación meteorológica	Santa Marta	87,72%	4
Estación meteorológica y mareográfica	Bahía Málaga	84,95%	4
Estación meteorológica y mareográfica	Quitavieja	18,55%	4
Estación mareográfica	Barranquilla	88,75%	2
Estación mareográfica	Isla Fuerte	89,25%	2
Estación mareográfica	Turbo	97,37%	2
Estación meteorológica	Punta Espada	99,61%	2
Estación meteorológica y mareográfica	Puerto Brisa	99,91%	2

²⁵ Fuente: Dimar-RedMpomm.

2.2.5. Etapa 5. Control de calidad de datos

Teniendo en cuenta que “el objetivo del control de calidad consiste en verificar si el valor de un dato notificado es representativo de la medición que se pretendía efectuar y no se ha visto afectado por factores no relacionados con el mismo” (OMM, 2011), en esta etapa se procedió con el control de calidad de los datos recuperados. Para ello, en primer lugar se definió el alcance de la calificación de los datos, para lo cual se seleccionaron tres tipos de banderas de calidad (QF, por sus siglas en inglés) de primer nivel recomendadas por el programa internacional IODE (Tabla 5).

Tabla 5. Extracto de banderas de calidad IODE aplicados por la RedMpomm²⁶.

Código QF	Significado	Descripción
2	Desconocida	No se cuenta con suficiente información para determinar la calidad del dato.
4	Malo	Los datos han fallado en una o más de las pruebas de control de calidad documentadas.
9	Dato faltante	Datos faltantes en la serie de datos.

A continuación, se tuvieron en cuenta las recomendaciones de IOOS (2014) para el control de calidad de datos de viento y las de IOOS (2016) para datos de nivel del mar, acordando y documentando cuatro pruebas de calidad básicas; estas son:

- **Prueba 1. Validación de sintaxis:** Se utiliza para descartar los mensajes recibidos que contengan una estructura desconocida o que el valor del parámetro o fecha de muestreo contenga caracteres no válidos. Como resultado, los datos son descartados automáticamente.
- **Prueba 2. Agregación temporal:** Esta prueba evalúa que el dato medido se encuentre dentro del periodo o ventana de tiempo esperada, es decir, de acuerdo a la agregación temporal configurada para cada variable en el sistema. Si se encuentran fuera del periodo de muestreo son descartados automáticamente.
- **Prueba 3. Datos faltantes:** En esta prueba se revisa toda la serie de datos en búsqueda de campos vacíos. Una vez identificados estos campos, se llenan con el valor “-99999” y se etiquetan con la bandera de calidad “9”.

²⁶ Fuente: Dimar-Cecoldo.

- **Prueba 4. Valores imposibles:** En esta prueba se aplica la recomendación de NOAA, 2015 citada en CTN Diocean (2018), la cual explica que “en general las pruebas de calidad deben incluir la identificación de valores físicamente improbables, que alcancen las variables en condiciones ambientales usuales”. Como resultado de esta prueba, se aplica la bandera “4” para los datos que se encuentran fuera del rango de medición propio del sensor (es decir, el valor establecido por el fabricante), y se asigna la bandera “2” para las demás mediciones.

Una vez aplicadas las pruebas de calidad en mención a cerca de 16 millones de datos de meteorología marina recuperados para las variables específicas de humedad relativa, temperatura del aire, velocidad y dirección del viento, y presión barométrica, para todas las estaciones en los periodos de tiempo disponibles, se obtuvieron los resultados que se aprecian en la Figura 15.

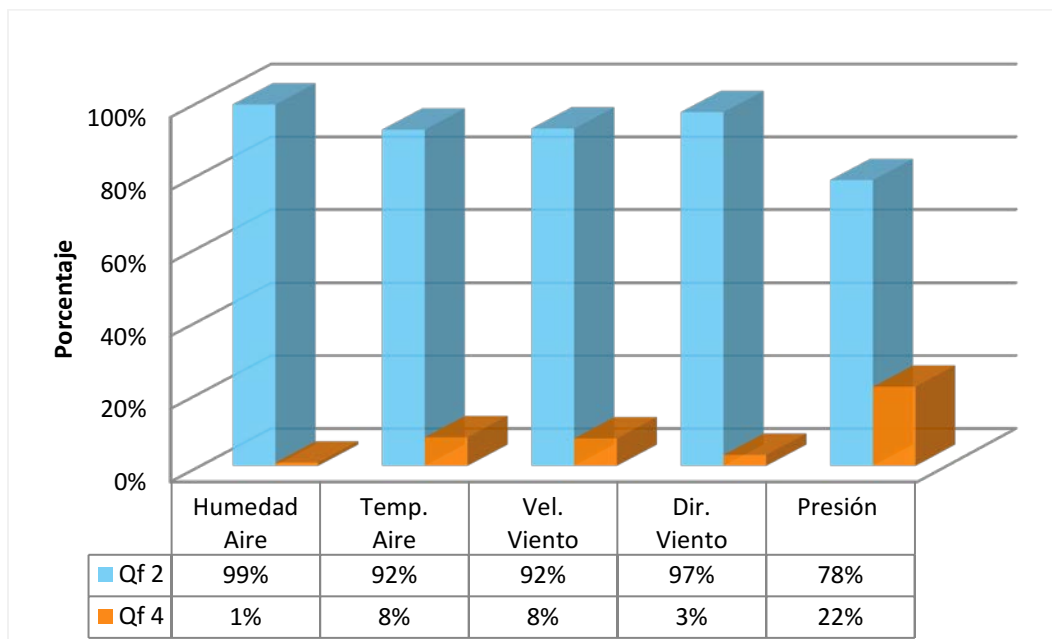


Figura 15. Resultado del proceso de calificación de datos de meteorología marina de la RedMpomm 2005-2018²⁷.

La mayor cantidad de datos calificados con QF “4” o datos “malos” correspondieron al parámetro de presión atmosférica, probablemente asociado al efecto de las condiciones del área de muestreo sobre el sensor de medición; para todos los demás parámetros se obtuvieron datos calificados con QF “2” por encima del 90%, es decir, datos que tendrán que ser sometidos a pruebas y análisis adicionales para determinar su calidad, en aplicación de la recomendación de OMM (2017) “El control de la calidad garantiza la calidad y coherencia de los datos obtenidos, que se consigue mediante un conjunto de procedimientos minuciosamente diseñados que se centran en las prácticas de buen mantenimiento, reparación, calibración, y verificaciones de la calidad de los datos”.

²⁷ Fuente: Dimar-RedMpomm.

2.2.6. Reflexiones finales

Valeh Nazemoff, vicepresidente senior y cofundador de Acolyst explica que “los datos son el activo más importante de un organización y, sin normas y sin calidad de datos, la organización no funciona”²⁸; por esta razón es importante desarrollar iniciativas para aumentar el valor de los datos, lo cual se logra documentando, formalizando y aplicando procedimientos, guías y buenas prácticas en todo el ciclo de vida del dato, desde su planeación hasta su almacenamiento, conservación y disposición; de esta manera el gobierno de datos no estará supeditado al criterio del administrador de turno.

Es importante tener en cuenta durante el diseño de cualquier sistema de información que incluya variables de monitoreo océano-atmosférico, la implementación de una base de datos “primaria” que contenga los datos tal cual como se capturaron, sin validaciones o correcciones, y sin más ajustes que los de su transformación de formato para su almacenamiento; lo anterior, para garantizar la trazabilidad de los datos. Asimismo, se deben establecer controles (físicos y lógicos) para mantener la integridad de dicha base de datos.

El Centro Colombiano de Datos Oceanográficos²⁹ contribuye con el cumplimiento de los objetivos del programa IODE de Unesco-COI en el nivel nacional, entre ellos se encuentra la gestión de mejores prácticas para el manejo, intercambio y acceso a datos; por ello es recomendable establecer comunicación con el Cecoldo durante la planeación de iniciativas de arqueología y recuperación de este tipo de datos.

La RedMpomm de Dimar tiene planeado desarrollar en el corto plazo la arqueología y recuperación de datos oceanográficos para el parámetro de oleaje. Se estima recuperar una serie de hasta 10 años de variables tales como altura, periodo, dirección y dispersión espectral del oleaje, para siete puntos de monitoreo ubicados en Tumaco, Isla Gorgona, Buenaventura, Bahía Solano, Turbo, Barranquilla y Puerto Bolívar.

²⁸ Recuperado de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/406201/qu-es-el-gobierno-de-datos-y-por-qu-lo-necesito>.

²⁹ Disponible en: <https://cecoldo.dimar.mil.co/>.

2.3. Proceso de preservación de la colección de aerofotografías históricas de Colombia: inventario, digitalización y disposición para consulta y desarrollo de estudios aplicados

Carantón y Álvarez (1998) definió una fotografía aérea o aerofotografía como la “imagen de un terreno captada desde un avión o cualquier otra nave aérea equipada con cámaras fotográficas especiales para tal fin. Es el registro completo y detallado, en el instante de la toma, de los elementos que conforman el medio ambiente geográfico, es decir, de los fenómenos físicos del medio natural y del complejo de acciones que el hombre realiza en ese medio”. Colombia es uno de los primeros países del continente en utilizar las fotografías aéreas para el levantamiento cartográfico de su territorio, siendo el IGAC el líder de esta tarea.

Desde sus inicios el Instituto Geográfico Agustín Codazzi (IGAC) fue concebido como un ente productor de cartografía, donde sus insumos principales a través del tiempo han sido las aerofotografías. Todo comenzó en 1902 cuando se creó un grupo de trabajo denominado en ese momento “Oficina de Longitudes y Fronteras” adscrita al Ministerio de Relaciones Exteriores, la cual sólo en 1910 comenzó a actuar, y doce años después inició labores de fotogrametría trabajando en conjunto con la empresa de aviación Sociedad Colombo Alemana de Transporte Aéreo (SCADTA), donde se llegaron a capturar sobre el Río Magdalena aproximadamente 700 aerofotografías. En 1935 se nombra como Instituto Geográfico Militar, y quince años después adopta el nombre que tiene actualmente en memoria del cartógrafo italiano Agustín Codazzi (IGAC, 2014).

Este proceso se ha desarrollado por más de 84 años; en un primer momento se capturaban imágenes del territorio colombiano con el uso de cámaras análogas, obteniendo películas de negativos, que se almacenaban por medio de rollos. En 1956 inició la construcción de la infraestructura que hoy en día se conoce, y en el 2007, 51 años después, el IGAC mediante el proyecto “Mejora de los sistemas de cartografía del territorio colombiano” y con financiación de recursos donados por la Comunidad Europea, adquirió un sistema de cámara aérea digital para la toma de aerofotografías digitales, además de la adquisición de tres escáneres fotogramétricos, con el objetivo de preservar el patrimonio histórico acumulado de rollos de negativos de película aérea (IGAC, 2014).

La colección histórica de rollos de negativos de películas aéreas, conformada desde 1935 hasta 2007, está integrada por 5.734 rollos, cada uno de los cuales contiene aproximadamente entre 200 y 270 fotografías en soporte plástico. En total la serie de imágenes reúne aproximadamente 1.400.000 aerofotografías monocromáticas en negativo (blanco y negro).

El soporte plástico de los negativos corresponde principalmente a acetato de celulosa, Politereftalato de Etileno (PET) y papel resina; y en menor medida nitrato de celulosa. La imagen de las fotografías se encuentra formada por haluro de plata aglutinada con gelatina.

Es importante resaltar que este tipo de fotografías tienen una gran fortaleza y una gran debilidad. La fortaleza recae en la estabilidad química y física de la técnica de la imagen, ya que los haluros de plata aglutinados con gelatina son muy estables al paso del tiempo. Si el soporte de la imagen es papel o papel resina, como en este caso, la fotografía es aún más estable, por la estructura química del papel.

La debilidad surge cuando el soporte es de plástico, bien sean de nitrato de celulosa, acetato de celulosa o PET; en este caso los plásticos sufren procesos autocatalíticos, emanando algunas de las sustancias durante los procesos de envejecimiento; dichos procesos no se pueden detener y si los niveles de emanación son bastantes altos, el proceso se acelera notoriamente. Por esta razón es muy importante asegurar las condiciones adecuadas de almacenamiento de toda esta colección, condición que ralentiza en gran medida el proceso, pero permite la preservación del material fotográfico por un largo tiempo.

Otro factor importante considerado en el proceso de preservación de la información de aerofotografías, es el riesgo tecnológico relacionado por una parte, con la limitada disponibilidad de escáneres fotogramétricos que permitan digitalizar las imágenes de los rollos de negativos de fotografía aérea en el IGAC, y por otra, el hecho de que estos equipos ya no se están siendo producidos, lo que lleva a que cada vez sea más difícil su consecución en el mercado y dificulta el mantenimiento durante su vida útil. Este riesgo implica para el IGAC un reto importante en cuanto a lograr la digitalización de los negativos de aerofotografía faltantes en el menor tiempo posible y garantizar de esta manera la preservación de la información recopilada que hace parte del patrimonio institucional.

Cabe mencionar que para salvaguardar la información de aerofotografías históricas del territorio nacional se requiere, de una parte, digitalizar el material, y de otra, garantizar su conservación en el formato original en el que fueron tomadas, es decir, en los rollos de negativos de película aérea, disponiendo las condiciones adecuadas para su almacenamiento, adelantando los procedimientos técnicos y administrativos para que este archivo sea declarado por el AGN como bien de interés cultural.

El proceso de preparación, manipulación, inventario y almacenamiento del material que será digitalizado, se efectúa bajo los lineamientos y asesoría técnica del AGN; básicamente consta de las etapas que se describen a continuación.

2.3.1. Preparación y almacenamiento

El material se encuentra ubicado actualmente en una bóveda del IGAC, construida hace 60 años aproximadamente. El ingreso a la bóveda se realiza a través de una puerta de seguridad metálica que permite tener mayor control en el acceso de personas, consecuente con el hecho que es información delicada de alta importancia nacional la que reposa allí (Figura 16).



Figura 16. Foto del cartel de la entrada de la bodega de rollos³⁰.

La bóveda es un recinto de dos pisos comprendida por una planoteca en el primer nivel (Figura 17) y en el segundo se almacenan las películas más antiguas de la colección. De manera organizada, en cada espacio se han dispuesto los empaques que son cajas cerradas tubulares con los rollos de negativos de película aérea, los cuales son manipulados para su inventario o digitalización (Figura 18).



Figura 17. Foto interna de la bodega (Planoteca)³¹.

³⁰ Fuente: Infraestructura Colombiana de Datos Espaciales (ICDE).

³¹ Fuente: Subdirección de Geografía y Cartografía del IGAC.

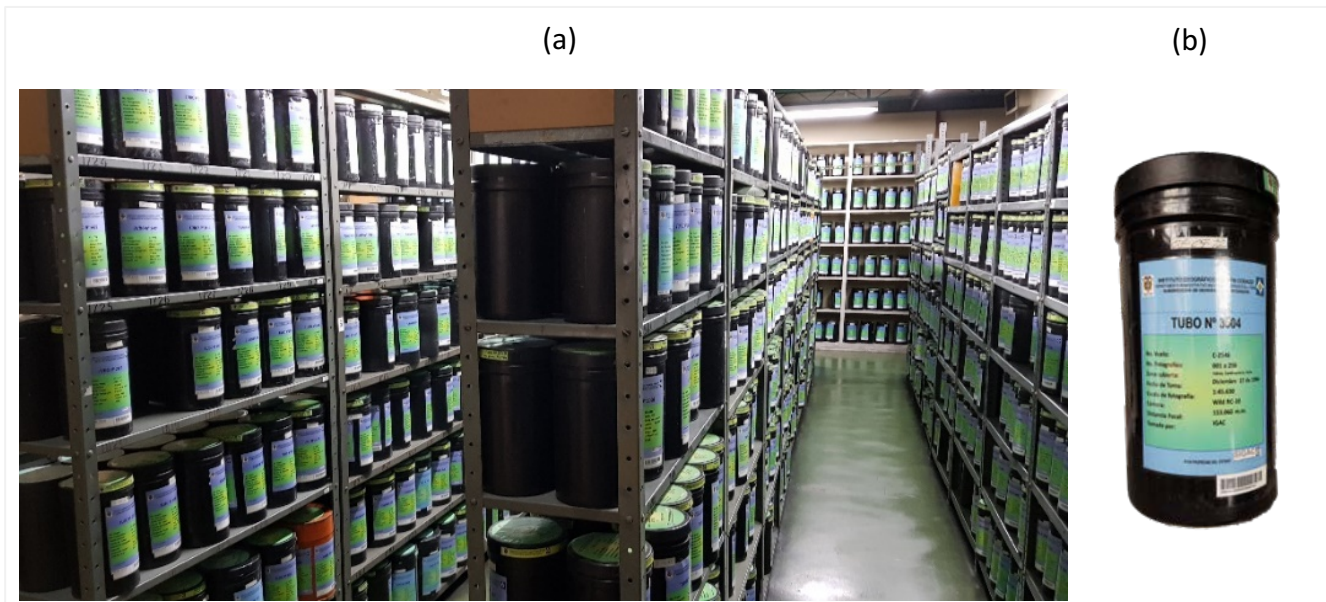


Figura 18. (a) Foto interna de la bodega con los estantes de rollos y (b) rollo fotográfico³².

En este recinto se ha realizado modificaciones que atienden las recomendaciones de curadores expertos en este tipo de preservación en cuanto al almacenamiento y manejo de las imágenes análogas, conforme a las directrices del AGN; por ejemplo, una de estas recomendaciones guarda relación con la habilitación y mantenimiento de los motores que absorben y extraen aire, buscando el flujo constante dentro de la bóveda de almacenaje; también dentro de las recomendaciones se encuentran las orientadas a que los estantes sean de material plástico y no metálico, y las pinturas no pueden ser de aceite; todo en función de la preservación de los rollos fotográficos.

En la actualidad se está generando un nuevo inventario bajo los parámetros establecidos por el AGN, para lo cual el IGAC ha dispuesto el personal requerido. Así mismo, se ha llevado a cabo la adecuación de una zona para la manipulación de rollos, donde se ubicaron varios extractores y unos repositorios con carbón vegetal que, junto con la aspersión diaria de una solución de cloruro de calcio disuelto en agua destilada, ayudan a la decantación de los materiales particulados que emanan los rollos al momento de retirarlos de sus recipientes contenedores.

Otro aspecto importante corresponde a la indumentaria obligatoria para la manipulación, en este caso se compone de gafas, máscaras especiales, guantes y batas blancas para los funcionarios que interactúan en estos espacios.

³² Fuente: Subdirección de Geografía y Cartografía del IGAC.

En la bóveda de almacenamiento, ubicada en el segundo piso del edificio principal del IGAC, se encuentra instalada una banda transportadora (Figura 19) a través de la cual descienden los rollos de manera mecánica al piso inferior, donde se localiza el área habilitada para la manipulación de los rollos y su inventario en el antiguo laboratorio fotográfico. En este espacio se encuentran los escáneres fotogramétricos para digitalización de las aerofotografías; también en esta zona, en varios de los antiguos cuartos oscuros, aún se encuentran equipos fotográficos, que por su aproximación a la obsolescencia técnica se dispondrán en el futuro en un museo cartográfico, previsto para conformar en dicha área.



Figura 19. Banda transportadora de rollos en la bodega³³.

2.3.2. Manipulación

Para realizar el inventario o digitalización los rollos, estos son transportados a través de la banda transportadora el día anterior al que van a ser intervenidos, colocándolos sobre una mesa recubierta con una tela especial; los rollos se mantienen destapados toda la noche para que expulsen los gases acumulados. Si llegasen a ser rollos muy antiguos, estos se dejan desde el viernes anterior a la semana en que van a ser manipulados, con el objetivo que durante el fin de semana se ventilen y se disipen los gases.

Adicionalmente, se tienen ventiladores constantemente encendidos para la circulación de aire; precisamente uno de los mayores agentes en contra de la preservación es el polvo, ante lo cual es importante manejar un ambiente lo más estéril posible. En varias esquinas del área de trabajo se colocan repositorios con carbón, para ayudar a la decantación de los materiales particulados, lo cual ayuda a mitigar los olores, reduciendo el riesgo para la salud de los funcionarios que realizan los procesos asociados al inventario o digitalización.

³³ Fuente: ICDE

2.3.3. Inventario

Un funcionario se encarga de verificar el estado de cada rollo y levantar los datos de las imágenes tales como vuelo, fecha, cantidad de escenas por rollo, entre otros (Figura 20); esta persona permanentemente utiliza máscara antigases con gafas, ya que los químicos que expelen los rollos son fuertes. Para la inspección de los rollos se emplea una máquina especial dispuesta para tal fin con vidrio a través del cual se proyecta luz tipo LED, además está dispuesta de manijas transportadoras de la película para inspección de cada fotografía por rollo. Simultáneamente se va diligenciando un formulario digital donde queda registrado el resultado de la inspección realizada.



Figura 20. Manipulación de rollos³⁴.

2.3.4. Digitalización

El proceso de digitalización se efectúa con escáner fotogramétrico; dependiendo de su tipo puede realizarse la digitalización de forma manual o automática. En el caso del escáner automático, el equipo utiliza encendido permanente, debido a que la digitalización puede tomar hasta tres días por rollo, funcionando 24 horas. Después de esto y teniendo las imágenes digitalizadas, mediante un programa de software se realiza su organización y clasificación.

La digitalización se lleva a cabo a partir de dos frentes de trabajo:

- **Plan de trabajo.** Hace referencia al plan de digitalización de rollos completos de negativos de películas. En este proceso se hace uso de los escáneres automáticos (Figura 21) y se lleva un registro de control de las aerofotografías dispuestas en formato digital.

³⁴ Fuente: Subdirección de Geografía y Cartografía del IGAC.



Figura 21. Sala de digitalización con escáneres automáticos³⁵.

- **A demanda.** Corresponde a la atención de requerimientos específicos de usuarios que solicitan un número limitado de imágenes a partir de la consulta del archivo aerofotográfico. Para esta labor se emplea un escáner manual.

Cabe aclarar que las imágenes en su mayoría están disponibles sin georreferenciar, ya que esta es una de las actividades que adelantará el IGAC como parte del proceso de preservación de la información.

2.3.5. Documentación

Actualmente se documentan manualmente los metadatos de las imágenes de aerofotografías bajo un perfil del estándar ISO 19115, utilizando el software de catálogo con el que cuenta el IGAC para tal fin.

2.3.6. Disposición y uso

Después de digitalizadas, las imágenes son almacenadas y dispuestas en el Banco Nacional de Imágenes (BNI³⁶), una solución desarrollada por el IGAC para permitir la consulta en línea de las imágenes de sensores remotos, así como de otros productos derivados con los que cuenta el Instituto.

³⁵ Fuente: ICDE.

³⁶ Disponible en: <http://www.bni.gov.co/>

2.3.7. Reflexiones finales

A la fecha se han digitalizado cerca de 331.800 imágenes correspondiente a un avance aproximado del 23,7% de más de 1.400.000 disponibles, que como se mencionó, están contenidas en la totalidad de los rollos almacenados en el IGAC. Teniendo en cuenta que el tratamiento de cada aerofotografía toma un tiempo cercano a 23 minutos y el cuidado que se debe tener con el manejo del material y en especial los rollos más antiguos, se estima que la tarea de digitalizar los rollos restantes requerirá de varios años más. Considerando el tiempo y la inversión requerida, el IGAC está gestionando recursos y alianzas interinstitucionales para que otras entidades coadyuven en la preservación del archivo aerofotográfico de la nación, en la búsqueda de aunar esfuerzos para disponer esta importante información para diferentes usos.

Se destaca la importancia de estos insumos para la nación, al ser el único archivo de esta naturaleza en el país, compuesto por información invaluable para numerosos estudios y una referencia histórica del territorio colombiano. También puede ser considerado el punto de partida para conocer la dinámica geográfica que ha tenido el país, desde el punto de vista del cambio multitemporal en el uso y ocupación del territorio.

Finalmente, el fundamento del proceso preservación y conservación debe entenderse en su verdadero alcance, no es solo como un proceso de digitalización para salvaguardar la información aerofotográfica registrada en los rollos de película en medio digital, sino también para conservar este archivo como un bien histórico y cultural de la nación, reduciendo su deterioro y garantizando su adecuada preservación para las generaciones futuras.

2.4. Reseña de la arqueología y recuperación de información para el “Banco de Información Petrolera”

La realización de la arqueología y recuperación de información para el Banco de Información Petrolera (BIP) ha sido de gran importancia; desde sus inicios se han utilizado metodologías basadas en técnicas que hacen referencia al proceso sistemático de recuperación de datos grabados en medios obsoletos, así mismo, se ha ido asegurando la información almacenándola en sistemas informáticos que día a día se van haciendo más sofisticados tecnológicamente.

Tan diversa como lo es la información geo-científica, han sido los procesos y actores que han llevado al BIP a su estado actual en el Servicio Geológico Colombiano. En este esfuerzo han estado presente profesionales expertos en la producción, gestión, organización y uso de datos tales como doctores, geo-científicos, entre otros, y en los procesos se han considerado tópicos como la búsqueda, reconocimiento, recuperación, completitud y preservación de la información. A continuación, se presenta una reseña de dichos procesos.

2.4.1. El primer “Banco de datos”

En julio de 1983 nació en Ecopetrol la idea de organizar el primer “Banco de Datos” y para ello, se inició la búsqueda y clasificación de medios que se encontraban en diferentes lugares como el mismo Ecopetrol, donde sus funcionarios tenían bajo su responsabilidad información sísmica o de pozos grabada en dispositivos como cintas de 21, 9 y 7 *tracks* o pistas, cartuchos, entre otros (Figura 22); también se recurrió a la búsqueda de medios que podían estar en compañías operadoras, compañías de servicios, centros de procesamiento de datos sísmicos, etc.

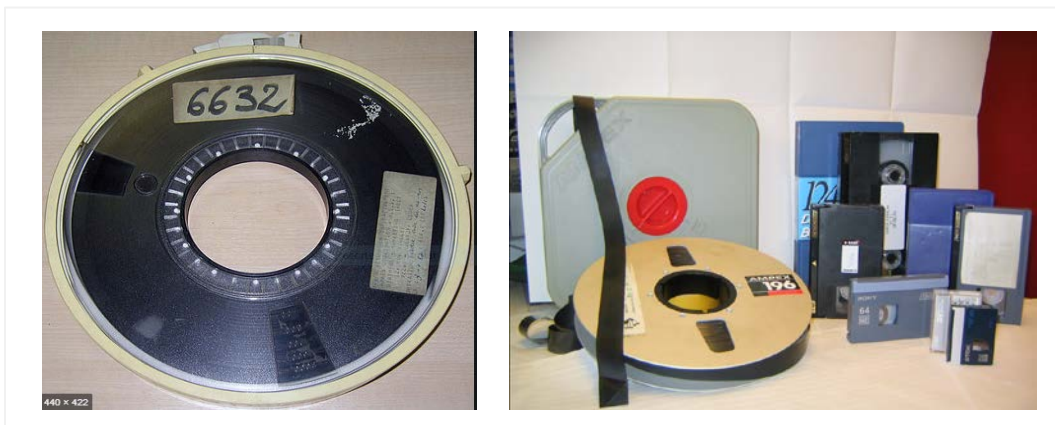


Figura 22. Tipo de medios desde donde se recuperaron datos para el BIP³⁷.

³⁷ Fuente: SGC.

Con este ejercicio, se logró reunir aproximadamente 65.000 medios magnéticos con datos sísmicos (primarios y procesados). Inicialmente, estos fueron almacenados en la ciudad de Bogotá donde se dio inicio a todo un proceso de gestión basado en la arqueología y recuperación de datos para que la información pudiera quedar disponible para toda la comunidad interesada.

Aunque ya se hablaba de un “Banco de Datos” desde 1983, fue hasta 1985 que inicia la historia como “Banco de Información Petrolera – BIP”, cuando Ecopetrol emprendió la labor de construir un espacio adecuado para conservar todos los medios recuperados en su mejor estado, y decidió hacerlo en la Vereda Mansilla del Municipio de Facatativá, Cundinamarca, Colombia, con la Cintoteca³⁸ NRP “Nelson Rodríguez Pinilla” como se conoce hoy en día (Figura 23).



Figura 23. Panorámica de la Cintoteca ubicada en el departamento de Cundinamarca³⁹.

Una vez terminada la construcción de dicha Cintoteca, se contó con espacios seguros y suficientes para que en 1987 se movieran desde Bogotá, los 65.000 medios magnéticos, así como medios análogos que representaba especialmente los reportes del observador, elevaciones, coordenadas, mapas y otros documentos, fruto de los levantamientos sísmicos que hasta ese momento se habían ejecutado en el país. Hecho lo anterior, en 1990 se continuó la arqueología de datos obteniendo información de los registros de pozos complementaria a la que ya estaba en la Cintoteca de sísmica, mapas y documentos técnicos. De esta manera, se logró contar con toda la información de E&P en un mismo lugar.

2.4.2. Desarrollo de una base de datos

En el año 1998, con toda la información bajo control en la Cintoteca, se evidenció la necesidad de llevar a otro estado, los medios e información análoga que allí se almacenaba. Entonces se emprendió una nueva iniciativa de arqueología y recuperación de datos que apuntó al sistema con el que se cuenta en la actualidad. Desde entonces y hasta el 2002, se aplicaron prácticas que permitieron preservar los medios magnéticos y análogos para llevar toda la información a medios modernos de grabación de datos como *Digital Line Tape* (DLT) y *Linear Tape-Open* (LTO) y en formato digital (Figura 24).

³⁸ En la Cintoteca NRP se preservan todos los medios físicos de información técnica de exploración y producción de hidrocarburos. Es allí donde se realiza el proceso de entrega de dichos medios por parte del EPIS (*Exploration and Production Information Service*) para su disposición final.

³⁹ Fuente: SGC.



Figura 24. Transformación de medios de almacenamiento de datos y documentos técnicos E&P⁴⁰.

Se desarrollaron programas que permitieron recuperar datos que estaban, inclusive, en formatos desconocidos que no correspondían a los manejados por la *Society of Exploration Geophysicists* (SEG). Los medios que para el momento presentaban hongos, se tornaron obsoletos, y para su tratamiento se aplicaron técnicas de horneado en hornos especializados, limpieza con equipos que retiraban las impurezas (sin dañar las cintas), y preservación basadas en el escaneo con Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés), lo que permitió al final desarrollar una base de datos desde donde la cual se podía tener acceso a toda la información dispuesta en formato digital.

En el 2003 el gobierno nacional decidió crear la Agencia Nacional de Hidrocarburos (ANH) y observó entonces la necesidad de transferir a esta entidad el BIP que había madurado bajo la administración de Ecopetrol; constituyéndose esta en la primera transferencia del BIP. En 2011 el gobierno nacional evaluó hacer una nueva transferencia del banco de información, en esta oportunidad al SGC, para lo cual estableció un plazo de ejecución de cinco años, desarrollándose de esta manera la operación del BIP en ANH hasta el 2 de noviembre de 2016.

2.4.3. El BIP hoy

Con motivo del traslado del BIP desde la ANH hacia el SGC, fue necesario recurrir nuevamente a técnicas de arqueología y recuperación de datos aplicados a medios de difícil lectura, ya sea por obsolescencia de los dispositivos de almacenamiento o por el mismo estado de los datos. Esta necesidad nació a partir de la identificación de información del BIP requerida por los usuarios, la cual, aunque ya se encontraba en repositorios digitales, al descargarla en su formato original resultaba dañada o encriptada. Para estos casos, se inició un proceso de búsqueda de medios antiguos que se almacenan en la Cintoteca NRP.

⁴⁰ Fuente: SGC.

Identificado el tipo de medio, se ubicaron en el mercado empresas especializadas que cuentan con lectores de medios antiguos, y con los controladores apropiadas con tasas de recuperación de datos y/o archivos por encima del 80%. Siguiendo esta estrategia, se encontraron proveedores con equipos disponibles para leer medios con datos de sísmica y pozos grabados, tales como cintas de 9 *tracks*, cartuchos 3480, 3480E, 3590, 3592, exabytes de 8 y 4 milímetros, entre otros.

En la actualidad, el SGC desarrolla arqueología y recuperación de datos en la medida en que estos son requeridos; la Cintoteca NRP es uno de los activos del BIP que almacena más de 1.3 millones de medios físicos, desde los cuales se puede recuperar la información solicitada, pudiéndose así atender la demanda de los usuarios.

La meta del BIP es lograr que toda la información con que cuente el banco de información sea lo más completa, confiable, íntegra y de la mejor calidad posible. Todo lo anterior, teniendo en cuenta que hoy por hoy se están desarrollando proyectos de tecnología enfocados a lograr una recuperación y completitud del 100% de la información para que esta quede disponible en línea.

3. PERSPECTIVAS ►

3.1. Arqueología y recuperación de registros biológicos de Parques Nacionales Naturales de Colombia

3.1.1. Generalidades

Parques Nacionales Naturales del Colombia (PNN) desarrolló y puso en operación el sistema de información de monitoreo de “Parques Nacionales Naturales de Colombia SULA” desde 2013, siendo esta una herramienta informática en línea⁴¹, bajo la arquitectura de Google App Engine, que apoya el proceso de estructuración y almacenamiento de datos de monitoreo, así como la visualización de los indicadores de estado, presión y respuesta de los Parques Nacionales Naturales de Colombia (PNN, 2018).

El sistema fue creado teniendo en cuenta herramientas de monitoreo, sin embargo desde el 2018 se agregó la temática de investigación científica, lo cual creó la necesidad de adelantar un proceso de arqueología y recuperación de información histórica sobre estudios puntuales o de línea base, incluyendo hábitat de ecosistemas terrestres y marino costeros, provenientes de caracterizaciones, informes de contratistas/consultorías, permisos o avales de investigación, artículos científicos, proyectos, recorridos de prevención, control y vigilancia. Todo lo anterior con el fin de integrar y disponer dicha información para PNN.

La información en SULA está organizada en carpetas temáticas, estructuras de datos y lotes de datos (Figura 25). En el caso de las investigaciones científicas, se integra a esta temática las siguientes estructuras de datos que se basan en el estándar DarwinCore que dispone el Sistema de Información de Biodiversidad de Colombia (SiB Colombia, 2019), y cuyas plantillas se adaptaron a las necesidades de organización de la información de investigación de PNN, así:

- **Listas de especies.** Son compilaciones de especies o taxones dentro de un conjunto de datos que puede tener un alcance temático o regional. Se usa para conformar las listas de especies registradas tanto en información primaria como en secundaria. Se organizan por grupos biológicos o por área protegida y no se incluyen coordenadas geográficas.
- **Registros biológicos** (observaciones y/u ocurrencias). Corresponden a avistamientos directos en campo de cualquier espécimen o a colectas temporales donde el individuo se libera en el mismo lugar que se encontró. Incluye el taxón y las coordenadas geográficas.

⁴¹ Disponible en: <http://sula.parquesnacionales.gov.co/>

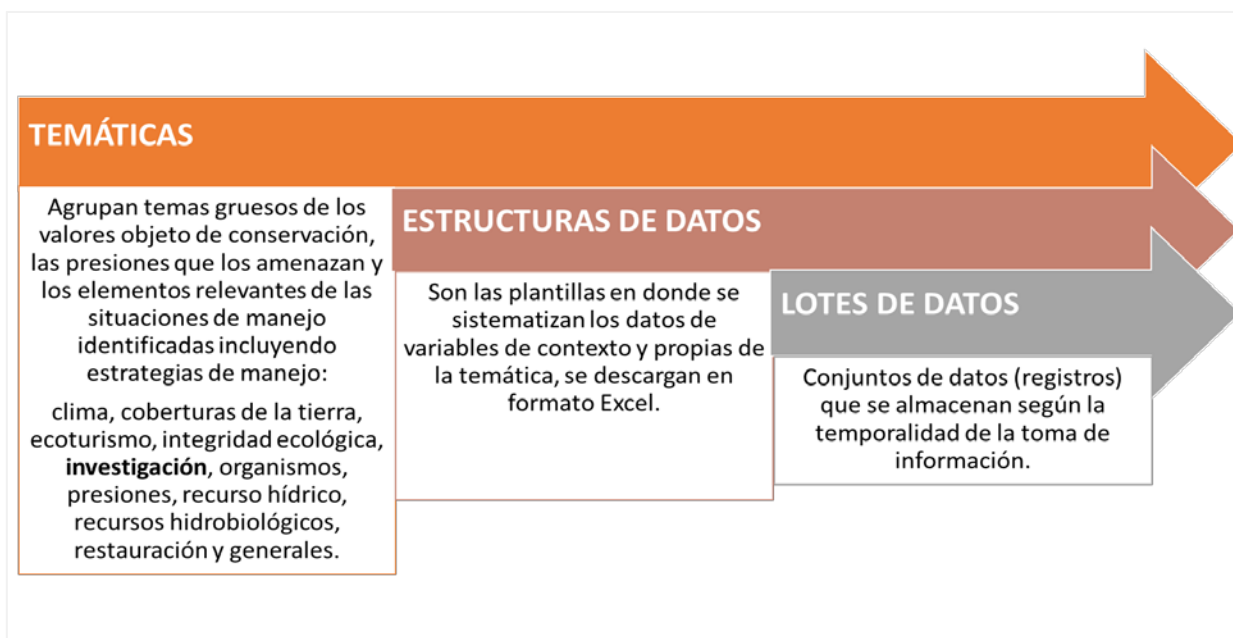


Figura 25. Carpetas de información en SULA⁴².

3.1.2. Avances alcanzados

Actualmente en la temática de investigaciones 16 áreas protegidas han vinculado información de 40 lotes de datos y 23.942 registros para los grupos biológicos que se aprecian en la Tabla 6. Constantemente se está ingresando nueva información usando como estrategias adicionales al trabajo de los equipos de las áreas, apoyo de guardaparques voluntarios y pasantes.

Tabla 6. Grupos biológicos integrados en la temática de investigaciones de SULA⁴³.

Estructura	Áreas protegidas	Grupo biológico	Lotes de datos	Registros
Lista de especies	Serranía de los Yariguies	Fauna y flora	2	1821
	Tinigua Central	Peces, mamíferos y plantas	3	305
		Aves	1	191
Registros biológicos	Churumbelos	Mamíferos	1	3505
	La Paya	Peces, aves, flora	6	852
	Chingaza	Herpetos	1	32
	El Cocuy	Stenocercus lache, Espeletia curialensis	2	346
	Estoraques	Entomofauna y plantas	2	6657
	Farallones	Herpetos	1	55
	Guacharos	Aves	1	48

⁴² Fuente: PNN

⁴³ Fuente: PNN

Estructura	Áreas protegidas	Grupo biológico	Lotes de datos	Registros
	Guanentá	Polylepis quadrujuga, Espeletia Congestiflora, Espeletia boyacensis, Espeletia incana	7	8306
	Los Flamencos	Aves	1	86
	Picachos	Peces	1	25
	Sanquianga	Peces	1	24
	Selva de Florencia	Herpetos	1	165
	Sumapaz			
	Tatama	Línea base frailejones	2	779
		Coleopteros, Libélulas, herpetos, orquídeas	4	737
	Tinigua	primates	1	8

La tarea es grande y corresponde a un primer aporte en la labor de sistematizar el conocimiento histórico base de las áreas protegidas del orden nacional. Se espera poder avanzar en el fortalecimiento de los centros de documentación de las áreas protegidas y las direcciones territoriales, iniciando con un inventario de la información tanto documental como de bases de datos, con el fin de tener un panorama claro del estado actual y poder avanzar en la gestión del conocimiento de la historia de los Parques Nacionales Naturales de Colombia.

3.2. Arqueología y recuperación de información para el mayor inventario geocientífico de Colombia

3.2.1. Generalidades

La información levantada en forma de productos por parte del Servicio Geológico Colombiano (SGC) es el alimento fundamental para nutrir los servicios de información geocientífica y nuclear de la entidad, que día a día suministra información a partir de solicitudes específicas, que buscan resolver problemas en las comunidades, en la industria y en el gobierno a nivel nacional y territorial. No obstante, la información producida no cumple con su propósito si no se facilita su consulta y si no es ampliamente divulgada.

Es por ello que el 28 de agosto de 2017 el SGC inició la operación del Motor de Integración de Información Geocientífica (MIIG)⁴⁴, concebido en la Arquitectura Empresarial para facilitar y apoyar los procesos misionales de generación y suministro de información a usuarios de diferentes sectores como son el científico, académico, público, privado y ciudadanía en general. El MIIG permite catalogar y organizar la información en repositorios centralizados que facilitan su acceso a los usuarios por medio de mecanismos de búsqueda, consulta y descarga del mayor inventario geocientífico de Colombia, compuesto por estudios misionales, archivos descargables y editables producidos en más de 100 años de producción de información.

El MIIG almacena aproximadamente 64.736 registros de metadatos que corresponden a 6.51 TB de archivos de los siguientes tipos: (i) 10.679 paquetes de estudios conformados por 21.716 documentos entre informes, memorias, anexos documentales, libretas de campo, figuras, etc.; y (ii) 32.341 mapas (cartografía, planchas, atlas, etc.).

Agrupado por periodos históricos, los paquetes de estudios disponibles en el MIIG son:

- 7997 producidos entre 1916 y 1999, que contienen 14.080 documentos y 22.956 mapas en formato PDF.
- 2682 producidos entre 2000 y 2019, que contienen 7.636 documentos en formato PDF y archivos editables con extensiones .DOC y .XLS, así como 9.385 mapas también en formato PDF y archivos nativos o editables en extensiones tales como .SHP, .COB, .E00, .MXD, .GDB, entre otros, de acuerdo a la evolución de los formatos SIG.

Como ejemplo de la información histórica disponible en MIIG, en la Figura 26 se aprecian los resultados de una búsqueda realizada en la plataforma, referente a documentos con información sobre el proyecto “Canal interoceánico de Napipi” generado en los años 20.

⁴⁴ Disponible en: <https://miig.sgc.gov.co>

The screenshot shows the MIIG platform search interface. At the top, there are search options: 'Búsqueda General', 'Búsqueda Avanzada', and 'Búsqueda Predefinida'. On the right, there are links for 'Acerca del MIIG' and 'Ayuda'. The left sidebar contains filters for 'Área del Conocimiento' (Geociencias Básicas: 5, Amenazas Geológicas: 1), 'Línea Temática' (Geología Económica: 3, Geología y Geomorfología: 2, Movimientos en Masa: 1), 'Escala' (Sin Escala Definida: 5, 1:1.500.000: 1), and 'Fecha' (a bar chart showing results from 1900 to 2019). The main area features a map of Central and South America with a search box containing the coordinates '-59.112, 9.237'. Below the map, there is a search bar with the text 'canal interoceánico' and a magnifying glass icon. Underneath, there are tabs for 'Mapas y Documentos' and 'Otros Recursos de Información'. The search results list three documents related to inter-oceanic canal projects in Colombia, with dates ranging from 1925 to 1929.

Figura 26. Aspecto de la plataforma MIIG para la búsqueda de información geocientífica del país⁴⁵.

3.2.2. Avances y metas

Estar atentos a los procesos de arqueología y recuperación de datos frente a los cambios tecnológicos y a las necesidades de preservación de la información, es una ardua labor que evita la pérdida de estos importantes activos para la nación. Implica el uso tecnologías de punta con parámetros bien definidos de acopio, respaldo y seguridad, que se logran con la participación de un equipo de profesionales comprometidos con estas actividades.

⁴⁵ Fuente: SGC

Por esta razón, desde el 2018 el SGC viene desarrollando una iniciativa de arqueología y recuperación de información, que incluye el mejoramiento del contenido de los metadatos y de los datos e información existentes en el inventario del MIIG, teniendo en cuenta parámetros adecuados de calidad, completitud y visualización de documentos, anexos y mapas que acceden los usuarios a través de esta herramienta.

Las actividades de arqueología y recuperación se han centrado en:

- Revisión detallada de elementos documentales y cartográficos listados en tablas de contenidos de los documentos principales.
- Búsqueda en diferentes fondos documentales en soporte físico, tales como documentos originales, mapoteca, biblioteca, en los cuales reposa la memoria misional histórica institucional.
- Digitalización de información generada entre 1916 y 1999, utilizando tecnología OCR y el formato PDF (Figura 27), que permita recuperar contenidos por medio del buscador de la plataforma MIIG.

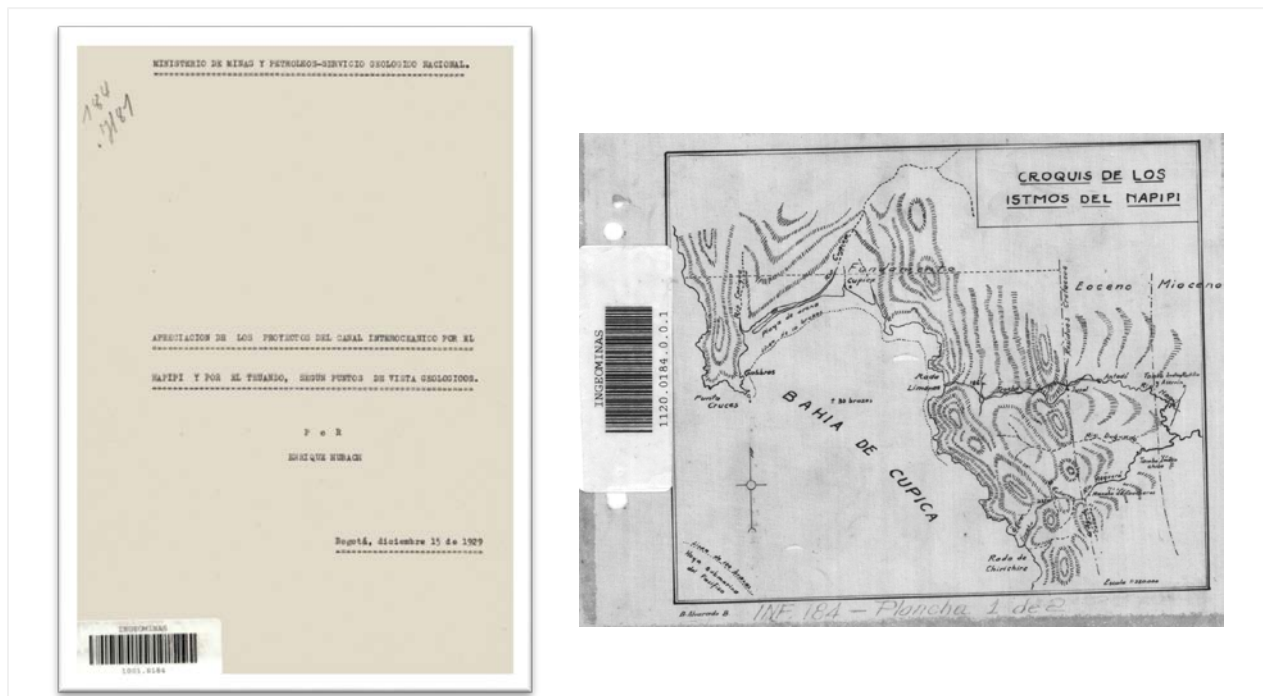


Figura 27. Documento producido en 1929 escaneado con funcionalidad OCR para facilitar su recuperación y mapa anexo disponibles en formato PDF⁴⁶.

⁴⁶ Fuente: SGC.

- Optimización del proceso de organización de la información y sus respectivos archivos nativos o editables.
- Catalogación de la información en áreas del conocimiento tales como geociencias básicas, recursos minerales, amenazas geológicas, hidrocarburos, asuntos nucleares y laboratorios, así como en líneas de investigación en geología, geomorfología, hidrogeología, geotermia, minerales metálicos, no metálicos y energéticos, recursos hidrocarburíferos, entre otras. Lo anterior con el fin de facilitar la búsqueda de la información a partir de términos o palabras claves, la aplicación de operadores booleanos, y mediante la interacción con un mapa – visor geográfico (búsqueda de los recursos que en sus valores de atributos coincidan total o parcialmente con las coordenadas dadas).

Como resultado de la implementación de las actividades en mención, en la actualidad se tiene un avance del 22% respecto al inventario total de registros de información documental y cartográfica, equivalente a 64.736 archivos revisados y actualizados. Se planea llegar a cubrir el 50% de la información a finales de 2020 y para el 2022 haber culminado el proceso de revisión de la calidad de datos, metadatos e información geocientífica del SGC.

4. BIBLIOGRAFÍA ▶

Andrade, C.A., Rangel, O.E. y Herrera, E. (2015). *Atlas de los Datos Oceanográficos de Colombia 1922-2013 Temperatura, Salinidad, Densidad, Velocidad Geostrofica*. Bogotá, Colombia: Dirección General Marítima-Ecopetrol S.A. Ed. Dimar. Recuperado de <http://cecoldodigital.dimar.mil.co/49>.

Carantón, D. y Álvarez, V. J. (1998). *Aplicación de las fotografías aéreas en geografía*. Bogotá, Colombia: Instituto Geográfico Agustín Codazzi. <http://biblioteca.igac.gov.co/janium/Documentos/1-00305.pdf>.

CCCP. (2004). *Proyecto "Central de Información Marina Colombiana, CENIMARC"*. Tumaco, Colombia: Dirección General Marítima.

CIOH. (2002). *Proyecto "Ampliación de los servicios de la Central de Datos del CIOH" Fase III/IV*. Cartagena, Colombia: Dirección General Marítima.

CIOH. (2005). *Proyecto "Recuperación y control de calidad de datos oceanográficos CIOH – Retrocean" Fase I/II*. Cartagena, Colombia: Dirección General Marítima.

Convenio 5211792 de cooperación para el desarrollo de actividades específicas y tecnológicas entre el Ministerio de Defensa Nacional – Dirección General Marítima, Ecopetrol S.A. y ACAC. Bogotá D.C., Colombia. 26 de julio de 2013.

CTN Diocean. (2018). *Manual de Referencia 2018 en Mejores Prácticas de Gestión de Datos Oceánicos*. Bogotá, Colombia: Dirección General Marítima. http://cecoldodigital.dimar.mil.co/2317/1/cc0_2539-2212_manual_referencia_mpgdo_3_2018.pdf.

Dimar. (1973). *Proyecto "Centro Colombiano de Datos Oceanográficos"*. Bogotá, Colombia: Dirección General Marítima y Portuaria.

Dimar. (1974). *Informe Crucero Oceanográfico Océano III - 1974*. Bogotá, Colombia: Dirección General Marítima y Portuaria.

Dimar. (1975). *Informe Crucero Oceanográfico Océano IV Áreas 2 y 3 - 1974*. Bogotá, Colombia: Dirección General Marítima y Portuaria.

Dimar. (1992). *Proyecto "Reactivación Centro de Datos Oceanográficos de Colombia"*. Bogotá, Colombia: Dirección General Marítima.

Dimar. (2016). *G1-GUI-001 "Guía para transferencias y disposición final documental", versión 0*. Bogotá, Colombia: Dirección General Marítima.

- Dimar. (2017). *M8-GUI-001 "Guía normalización de conjuntos de datos oceanográficos", versión 4*. Bogotá, Colombia: Dirección General Marítima.
- Dimar. (2018). *Informe Anual del Centro Colombiano de Datos Oceanográficos (Cecoldo) 2017*. Bogotá, Colombia: Dirección General Marítima. <http://cecoldodigital.dimar.mil.co/2184/>
- DMI. *Dublin core Metadata Initiative*. (Revisado el 16 de octubre de 2019). <https://www.dublincore.org/>
- Eco, H. (1984). *El Nombre de la Rosa*. Barcelona: Lumen.
- Hernández Jaimes, J. L.; Ortiz Martínez, R. V. y Suárez Pinzón, I. (2007). *Metodología archivística para la recuperación de información oceanográfica del Pacífico colombiano*. Boletín Científico CCCP (14). pp. 123-150. https://doi.org/10.26640/01213423.14.123_150
- IGAC. (2014). *IGAC 80 años*. (Revisado el 25 de noviembre de 2019). <https://www.igac.gov.co/es/contenido/igac-80-anos>.
- IOC. 1999. *International Global Oceanographic Data Archaeology and Rescue (GODAR) Review Conference, Silver Spring, MD, USA, 12-15 July 1999*. Paris, Intergovernmental Oceanographic Commission of Unesco. https://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=954
- IOOS. (2014). *Manual for Real-Time Quality Control of Wind Data: a guide to quality control and quality assurance for coastal and oceanic wind observations. Version 1.0*. <https://www.oceanbestpractices.net/handle/11329/264>.
- IOOS. (2016). *Manual for Real-Time Quality Control of Water Level Data: a Guide to Quality Control and Quality Assurance for Water Level Observations. Version 2.0*. <https://www.oceanbestpractices.net/handle/11329/267>.
- Méndez Rodríguez, E. (2002). *Metadatos y recuperación de información estándares, problemas y aplicabilidad en bibliotecas digitales*. Barcelona, Ediciones Trea.
- Méndez Rodríguez, E. (2003). *Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. *Anales de Documentación*, 295. <https://doi-org.ezproxy.javeriana.edu.co/10.6018/analesdoc.6.0.3591>
- MinTIC. (2014). *G.INF.06 Guía Técnica - Gobierno del dato. Versión 1.0 del 30 de diciembre de 2014*. Colombia: Ministerio de Tecnologías de la Información y Comunicaciones. https://www.mintic.gov.co/arquitecturati/630/articles-9258_recurso_pdf.pdf.

NISO. (2019). *Implementing Linked Library Data*. (Revisado el 10 de Octubreoctubre de 2019). <https://www.niso.org/events/2019/11/implementing-linked-library-data>.

NOAA. (2015). *Contents of World Ocean Atlas 2001 CD-ROMS* (Revisado el 15 de Octubreoctubre de 2019). <https://www.nodc.noaa.gov/OC5/WOA01/cdwoa01.html>.

OMM. (2011). *Guía de prácticas climatológicas OMM-N° 100. Edición 2011*. Ginebra, Suiza: Organización Meteorológica Mundial. http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf.

OMM. (2017). *Guía de instrumentos y métodos de observación meteorológicos OMM-N° 8. Edición 2017*. Ginebra, Suiza: Organización Meteorológica Mundial. https://library.wmo.int/doc_num.php?explnum_id=3664.

Ortiz Martínez, R. y Rico Lugo, H. (2006). *Herramienta Software para la Administración de la Información Marina y Costera Colombiana*. Boletín Científico CCCP (13). pp. 3-16. ISSN 0121-3423. https://doi.org/10.26640/01213423.13.3_16.

PNN. (2018). *Instructivo para la administración y manejo de la información de monitoreo*. <https://storage.googleapis.com/sula-files/help/es.pdf>.

SiB Colombia. (2019). *Wiki SiB Colombia*. <https://sites.google.com/humboldt.org.co/wikisib/publicar/est%C3%A1ndares>

Taylor, A. (2009). *The organization of information*. Westport, Connecticut: Libraries Unlimited.

5. SIGLAS

ACAC	Asociación Colombiana para el Avance de la Ciencia
AGN	Archivo General de la Nación
ANH	Agencia Nacional de Hidrocarburos
ARC	Armada de la República de Colombia
BIP	Banco de Información Petrolera
BNI	Banco Nacional de Imágenes
CECOLDO	Centro Colombiano de Datos Oceanográficos
CENIMARC	Central de Información Marina Colombiana
CCO	Comisión Colombiana del Océano
CCCP	Centro Control Contaminación del Pacífico
Cccp	Centro de Investigaciones Oceanográficas e Hidrográficas del Pacífico
CICAR	Programa de Investigaciones Cooperativas del Caribe y Regiones Adyacentes.
CIMI	<i>Consortium for the Interchange of Museum Information</i>
CIOH	Centro de Investigaciones Oceanográficas e Hidrográficas
Cioh	Centro de Investigaciones Oceanográficas e Hidrográficas del Caribe
COI	Comisión Oceanográfica Intergubernamental
CPC	Cuenca Pacífica Colombiana
CTD	<i>Conductivity, Temperature, and Depth</i>
CTN Diocean	Comité Técnico Nacional de Coordinación de Datos e Información Oceánica
DANE	Departamento Administrativo Nacional de Estadística
DC	<i>Dublin Core</i>
DCIM	<i>Dublin Core Metadata Initiative</i>
DIMAR	Dirección General Marítima
DLO	<i>Document Library Object</i>
DLT	<i>Digital Line Tape</i>
DOI	<i>Digital Object Identifier</i>
EAD	<i>Encoded Archival Description</i>
EPIS	<i>Exploration and Production Information Service</i>
ERFEN	Estudio Regional del Fenómeno El Niño/La Niña
E&P	Exploración y Producción
GILS	<i>Government Information Locator Service</i>
GODAR	<i>Global Oceanographic Data Archaeology and Rescue</i>
ICDE	Infraestructura Colombiana de Datos Espaciales

IGAC	Instituto Geográfico Agustín Codazzi
Invermar	Instituto de Investigaciones Marinas y Costeras
IODE	<i>International Oceanographic Data and Information Exchange</i>
IOOS	<i>Integrated Ocean Observing System</i>
ISBD	Descripción Bibliográfica Internacional Normalizada
ISBN	<i>International Standard Book Number</i>
ISSN	<i>International Standard Serial Number</i>
LTO	<i>Linear Tape-Open</i>
MARC	<i>Machine Readable Cataloging</i>
MEDI	<i>Marine Environmental Data Information Referral Catalogue</i>
MIIG	Motor de Integración de Información Geocientífica
NOAA	<i>National Oceanic and Atmospheric Administration</i>
NODC	<i>National Oceanographic Data Center</i>
OCLC	<i>Online Computer Library Center</i>
OCR	<i>Optical Character Recognition</i>
ODV	<i>Ocean Data View</i>
OEA	Organización de Estados Americanos
OMM	Organización Meteorológica Mundial
PET	Politereftalato de Etileno
PNN	Parques Nacionales Naturales de Colombia
QF	<i>Quality Flag</i>
RedMpomm	Red de Medición de Parámetros Oceanográficos y de Meteorología Marina
SCADTA	Sociedad Colombo Alemana de Transporte Aéreo
SEG	<i>Society of Exploration Geophysicists</i>
SGC	Servicio Geológico Colombiano
SiB	Sistema de Información de Biodiversidad
SPOA	Sistema de Pronósticos Oceánicos y Atmosféricos
URI	<i>Uniform Resource Identifier</i>
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
W3C	<i>World Wide Web Consortium</i>
W3CDTF	<i>Consortio Web, Data and Time Format</i>



**Ministerio de Defensa Nacional
Dirección General Marítima**

Carrera 54 No. 26-50 CAN. Edificio Dimar
Línea de Atención al Ciudadano 01 8000 115 966
Contact Center +57 (1) 328 6800 en Bogotá D.C.
Línea Anticorrupción 01 8000 911 670
dimar@dimar.mil.co

www.dimar.mil.co



Dirección
General Marítima



@dimarcolombia



DimarColombia



dimarcolombia



dimarcolombia
www.issuu.com



App Gente de Mar

