Dec 2019 / No. 4 Bogotá DC, Colombia ISSN 2539-2212 (Online)

Reference Manual

in Best Practices for Ocean Data Management

Archeology and recovery of data and information

2019

Page 2

C REVENUES

Best Reference Manual
Ocean Data Management Practices
Issue 4/2019

A digital publication of the Technical Committee
National Data Coordination
and Oceanic Information - CTN Diocean,
of the Colombian Ocean Commission - CCO
Bogotá DC, Colombia www.cco.gov.co

D IRECTION

Vice Admiral Juan Manuel Soltau Ospina
Maritime General Director - DIMAR

Captain Juan Camilo Forero Hauzeur
CCO Executive Secretary

C ONTAINED

Darío Fernando León Rojas
Ministry of National Defense - MDN

Ruby Viviana Ortiz Martínez
Wilder Antonio Álvarez Sánchez
General Maritime Directorate - Dimar

Martín Hernando González Martínez
Nancy Liliana Barreto Parra
John Guibsson García Guerrero
Fredy Alberto Gutiérrez García
Ana Alexandra Morales Escobar
Agustín Codazzi Geographical Institute - IGAC

Adriana Margarita Rozo Melo
National Natural Parks of Colombia - PNN

Pedro Augusto Rangel Segura
Rigoberto Blandón Grajales
Omar Gonzalo Santafé Alfonso
Jaime Alberto Garzón Barrios
Colombian Geological Service - SGC

R EVIEW OF E STILE AND C ONTENT

Ruby Viviana Ortiz Martínez
General Maritime Directorate - Dimar

Carolina Garcia
Institute of Marine and Coastal Research - Invemar

Nestor Ricardo Bernal Suárez
University Francisco Jose de Calda

C ONCEPT G RAPH

Strategic Communications Area (Acoes) - Dimar

E DITORIAL D IMAR

Photography:
Sampling with rosette. Oceanographic Cruise
Colombian Pacific Basin LIV / XVIII Regional Cruise
CPPS Ensemble, January 2016. Courtesy Dimar

ISSN 2539-2212 (online)

Page 3

Reference Manual on Best Practices for Ocean Data Management

Issue 4/2019

Bogotá DC, Colombia. December 2019

The "Reference Manual on Best Practices for Ocean Data Management" is a product of the Working Group on Best Data Management Practices (GT MPGD), which is part of the internal structure of the National Technical Committee Coordination of Oceanic Data and Information (CTN Diocean) of the Colombian Ocean Commission (CCO), and has Aim to recommend standards and best practices for the management of ocean data and information of the country in all its life cycle, taking into account national, regional and local levels.

It is technical and informative, in Spanish and in electronic format. The information and concepts expressed in this publication they must be used by the interested parties under their responsibility and discretion, however, it is understood that any divergence with what is published is of interest to the CTN Diocean, for which the sending of their Comments or suggestions to the email oceano@cco.gov.co .

Cite the complete work as follows:

CTN Diocean. (2019). *Reference Manual on Best Practices for Ocean Data Management. Number 4/2019* . Bogota DC, Colombia: DIMAR. DOIhttps://doi.org/10.26640/25392212.4.2019 .

Cite the chapters like this:

León Rojas, Darío Fernando. (2019). *Importance of metadata in information retrieval* . (Ed.), Manual of Reference in Best Practices for Ocean Data Management. Number 4/2019 (pp. 07-15). Bogotá, DC, Colombia: DIMAR.

Ortiz Martínez, Ruby Viviana. (2019). *Archeology and recovery of data collected over 40 years on board cruise ships Colombian oceanographic* . (Ed.), Reference Manual in Best Practices for Ocean Data Management. Number 4/2019 (pp. 16-26). Bogotá, DC, Colombia: DIMAR.

Álvarez Sánchez, Wilder Antonio. (2019). *Challenges in archeology and data recovery from the monitoring network largest meteomarino in Colombia* . (Ed.), Manual of Reference in Best Practices for Ocean Data Management. Number 4/2019 (pp. 27-36). Bogotá, DC, Colombia: DIMAR.

González Martínez, Martín Hernando; Barreto Parra, Nancy Liliana; García Guerrero, John Guibsson; Gutiérrez García, Fredy Alberto and Morales Escobar, Ana Alexandra. (2019). *Process of preservation of the collection of historical aerial photographs of Colombia: inventory, digitization and provision for consultation and development of applied studies* . (Ed.), Reference Manual in Best Practices for Ocean Data Management. Number 4/2019 (pp.37-44). Bogotá, DC, Colombia: DIMAR.

Rangel Segura, Pedro Augusto and Blandón Grajales, Rigoberto. (2019). *Overview of archeology and information retrieval for the "Petroleum Information Bank"* . (Ed.), Manual of Reference in Best Practices for Ocean Data Management. Number 4/2019 (pp.45-48). Bogotá, DC, Colombia: DIMAR.

Rozo Melo, Adriana Margarita. (2019). *Archeology and recovery of biological records of National Parks Natives of Colombia* . (Ed.), Best Reference Manual Ocean Data Management Practices. Number 4/2019 (pp.49-51). Bogotá, DC, Colombia: DIMAR.

Santafé Alfonso, Omar Gonzalo and Garzón Barrios, Jaime Alberto. (2019). *Archeology and information retrieval for the elderly geoscientific inventory of Colombia* . (Ed.), Reference Manual in Best Practices for Ocean Data Management. Number 4/2019 (pp.52-55). Bogotá, DC, Colombia: DIMAR.

Page 4

C ONTENT

P.

National Technical Committee for Coordination of
Ocean Data and Information

LIST OF FIGURES

P.

TABLE LIST

P.

INTRODUCTION

The National Technical Committee for Ocean Data and Information Coordination (CTN Diocean) is allows to present the fourth number of its "Reference Manual on Best Data Practices Oceanic ", which this time addresses the" archeology and recovery "of data and information. This theme, in the context of data management and more specifically of the *Global* initiative *Oceanographic Data Archeology and Rescue* (GODAR) refers to a two-stage process: the first " *archeology* ", which consists of searching, identifying, evaluating and describing historical data so that be useful to the community; and the second " *rescue* " refers to the effort to store and preserve the data, by digitizing, copying on electronic media, and filing in databases or digital repositories (Adapted from IOC, 1999).

Through this publication, the member institutions and universities of the CTN Diocean, give know the effort that Colombia has been making in the field of "archeology and recovery", to increase historical and recent archives of ocean data and information in digital format, with the purpose of disposing them to the community in a standardized and interoperable way for reuse.

In the first chapter reflections are raised about the importance of metadata in the information retrieval, seen from the cataloging perspective, that is, from the structuring of the data from documenting the greatest amount of information associated with it, but also from the point of view of the contributions made by the metadata in the search processes data and information on the internet.

The second chapter brings together experiences in archeology and data recovery and information in different disciplines; The first of these correspond to two initiatives led by the General Maritime Directorate (Dimar) through which part has been recovered, the historical heritage of oceanographic and marine meteorology data obtained in more than one hundred of oceanographic cruises developed in Colombia; on the other hand, the time series of metocean data whose measurements in real time at different points of the Caribbean, Pacific coast and Colombian insular areas, began in 2005 and extend to the present day.

The Agustín Codazzi Geographical Institute (IGAC) explains the complexity of the preparation process, handling, inventory, storage, digitization and documentation of the collection of historical aerial photographs of Colombia, which has had the technical advice of the General Archive de la Nación (AGN), and how this effort has made it possible to have part of the material in digital format for citizen consultation.

For its part, the Colombian Geological Service (SGC) makes a quick review of the history of how the Petroleum Information Bank (BIP) has been structured in recent years, in relation to with the activities developed to retrieve and carry the technical information of Exploration and Production (E&P) to such an organization, which allows to have an easy and safe way to be used in the development of projects of national and international interest.

Finally, the third chapter presents the advances that both National Natural Parks of Colombia (PNN) and the SGC have obtained, carrying out archeology and recovery activities of data and information, once institutional platforms and software have come into operation in disciplines such as geosciences and biology. The above shows that the mentioned activities are present in any of the stages of the information systems and for common, which are to dispose and preserve data and historical and recent information with the best quality possible.

1. IMPORTANCE OF METADATA IN THE RECOVER OF INFORMATION

1.1. Background

All civilizations have produced data, information, and knowledge; to prevent these fall into the oblivion of orality, must be encapsulated in objects that can be stored and consulted; such objects can be documents, as exposed in the early twentieth century the French librarian Suzanne Briet “All objects may or may not become a document […] example an Antelope running on the wild plains of Africa is not a document, but it is captured and taken to a zoo or turned into an object of study yes it is ”.

Currently, documents are available in physical formats (generally on paper) or in digital formats that are consulted through electronic means, and as the Document generation is more challenging to manage. So in order to spread the written knowledge mechanisms must be created to achieve its easy location either physically or digital, which is why, from ancient empires to the present day, society has developed libraries or information systems to manage documents under procedures standardized capture, control and dissemination.

The first step to facilitate or restrict access to documents is to define the procedure of organization and distribution of the documentation, either in a physical medium through shelves or in a digital medium through a screen. Humberto Eco (1984) in his book "The name of the Rose" develops a detective plot around a book by Plato, which was stealthily guarded in the library of a medieval Abbey, with organization in the form of a labyrinth and also with means security as poison staining the sheets of books, to avoid what could currently approaching an "Information leak".

For this organization to allow people who did have access to Plato's book the consultation was achieved without major difficulty, the librarian had a document in which in order alphabetically organized the titles of each book stored in the library, and next to each book registered an acronym and a number that respectively represented the living room and the shelf where This was located (Figure 1). The mentioned document created and used by every library in the world for more than a thousand years it was known as a “Catalog” and the information it contained is called in the present as "Metadata". Currently, thanks to new information technologies, metadata allow directing towards the location of information in an increasingly efficient way.

Figure 1. Plan of the library of the Benedictine Abbey, which is the center of the plot in which the novel takes place "The Name of the Rose" [1].

1.2. What is metadata?

Regardless of whether your application is basic or advanced, metadata does the job of describing the information that, seen from the perspective of Méndez Rodríguez (2003), is defined as:

- Highly structured data that describes the information, content, quality, condition and other characteristics of the data.

- "Information on information" or "data on data".

[1] Source: Eco, H. (1984) The Name of the Rose. Barcelona: Lumen.

The latter was the initial definition described by Richard Myers (Quoted by Méndez Rodriguez, 2003) in the sixties of the 20th century, where he proposes that all that component that identify a document is considered as metadata. Another conception like the one exposed by NISO (2019), defines metadata as "an information structure that describes, explains, locates in another way; it makes retrieving information as easy as possible for use and management . " For their full use Bearman and Sochats (Quoted by Taylor, 2009), define that the metadata must meet certain basic organizational conditions which are: Registration, terms and conditions (access), structure, context, content and history of use of the resource.

In general, any structured form used to describe and organize information can be considered as metadata, as long as they fulfill the function of addressing a specific information, usually found inside a document. Inside of librarianship metadata is the basis of its bibliographic control and dissemination system; the latter in a principle was carried out by means of physical formats such as files or Description sheets International Bibliographic Standard (ISBD), but evolved thanks to technology, towards the storage in bibliographic databases, which through search engines allow to be human efficiency and precision in the location of information, either by providing points physical or digital access.

Likewise, in this new technological era, metadata allows the interoperability of information systems, with the help of transmission protocols such as Z39.50, making visible information automatically from different sources and separated by large distances, improving times for dissemination and communication of knowledge; this is like a "Great [electronic] information highway" (Méndez Rodríguez, 2002) standardized.

1.3. Metadata classification

Due to the wide variety of characteristics and uses of information, Méndez Rodríguez (2002) explains that there is metadata for:

- Describe information resources (objects), such as Dublin Core.
- Define the elements of the catalog records in libraries, within which is the *Machine Readable Cataloging* (MARC) format .
- The exchange of museum information, such as the *Consortium for the Interchange of model Museum Information* (CIMI).
- Describe geographic or geospatial resources, such as the ISO 19115 standard.
- Describe electronic government information, for example, the *Government* service *Information Locator Service* (GILS).
- For the archival description, such as the *Encoded Archival Description* (EAD) model.

Therefore, several authors simplify the classification of metadata into the following types, which generally depend on the needs of internal and external users:

- **Administrative** : They are metadata that facilitate the management and administration of objects or documents. These elements are used to control and maintain the quality of the object, define the current location and state of the object.

- **Descriptive** : They are the components that identify the resource. It is based on the basic elements description (eg Author, title, publisher, etc.) plus any special elements that are necessary to represent the object.

- **Conservation** : They are the ones in charge of the conservation of the information source as reference object. These elements are necessary due to the constant alteration that information suffers.

- **Technicians** : These metadata are responsible for storing information about the characteristics and needs of the object for its visualization.

- **Use** : These are metadata that are automatically generated by a code that calculates the use of the object. They are generally the accountants that report on the number of entries or downloads of a resource.

1.4. Recover of information

The "Semantic Web" is a set of technological tools for localization on the Internet which, through interaction with different library databases, digital repositories, information systems, etc., seek an adequate and standardized path for the recovery of information and thus cover a specific need. This requires data with content, meaning and preferably related to other data, thus avoiding events such as from Lewis Carroll's famous tale "Alice in Wonderland," where Alice asks the Cheshire cat at a certain point which way to go, and Cheshire replies: "That depends a lot of the place where you want to go. If you don't know where you want to go, no matter which way you go, either path will take you somewhere. ”

Metadata plays an important role in information retrieval, which must largely because these are considered "highly structured data"; means there is a rigor and structuring for each component of the metadata, which follows a low pattern segment or set specific parameters, thereby achieving homogeneity in the data stored to link to the resource. These parameters can be the following: semantics, kernel of metadata, boolean operators and truncators. Up next, each of the preview are described.

1.4.1. Semantics

According to the Royal Academy of the Spanish Language, semantics are the "aspects of meaning, meaning or interpretation of linguistic signs such as symbols, words, expressions or formal representations ”; It is the way in which each of the words or texts that is registered in metadata fulfills its function of representing what you want to show. It is important to keep in Account the following aspects of semantics in the process of creating metadata, so that the information search leads the user to the document or object of interest:

- ✓ Synonymy: refers to two words of similar meaning, but of different writing, p. ex. ship and ship.
- ✓ Polysemy: they are words with different meanings and similar writing, p. ex. saw which It can mean a geographical accident or a bias tool.
- ✓ Homonymy: they are words of reading tone, but with different writing, p. ex. wise and sap.

1.4.2. Metadata core

It is the format generally based on a standard, which defines the structure, the fields and the way to enter the texts that give the meaning to the metadata. As an example, below, describes the Dublin Core (DC) considered the most used standard for representation and retrieval of electronic information, due to its adaptability to changes in the internet and its Multidisciplinary application largely supported by the *Online Computer Library Center* cooperative (OCLC), and by *World Wide Web Consortium* (W3C).

Initially, DC was conceived as a self-generated description mechanism author of the document understood as *Document Library Object* (DLO), but its progress has made it in a “focus of attention of communities involved in the formal description of resources or units information (museums, libraries, government agencies and even commercial organizations) " (Méndez Rodríguez, 2002). For this reason, at present, it is a topic in constant development by part of the international information science communities.

The DC metadata core is made up of fifteen basic elements that can expand or omit depending on the needs of the object (Table 1). Unlike the structure MARC format, all DC components can be repeated as many times as necessary, due to the various states and structures that electronic information has.

Table 1. Elements of the Dublin Core [2] international standard.

DC element	Equivalence	Description and use
DC. Title	Title	Name given to a document usually by the author.
DC. Subject	Matter	Resource matters. Descriptors or phrases that describe the DLO content. The use of controlled vocabularies and information systems is encouraged classification or *schemes.*
DC. Description	Description	Description of the resource; a summary, in the case of a text document, or a description of the content in the case of a visual document.
DC Source	Source	Character sequence used to uniquely identify a job from from which the current resource comes.
DC. Language	Language	Language (s) of the intellectual content of the DLO. The content of this field must match ISO 639 (RFC 1766).
DC. Relationship	Relationship with others documents	Identifier of a resource and the relationship it has with the information object that is being described. This element allows you to link related resources and descriptions of resources. For example: *IsVersionOf* (Version of a work), *IsBasedOn* (translation based, fixes, etc.), *IsPartOf* (It is part of a major component), *IsFormatOf* (Transformation of a dataset).
DC. Coverage	Coverage	Characteristic of spatial and / or temporal coverage of the content of the resource. I know refers to the physical region […] generally it is data extracted from a list controlled or *scheme* . Temporary coverage refers to the temporality of the object.
DC. Creator	Creator, Author	Natural or legal person responsible for creating the intellectual content of the resource. For example, authors, in the case of written documents, and artists, photographers and illustrators, in the case of visual resources.
DC. Publisher	Editor	Entity responsible for making the resource available in its format current. (eg, publishing company, university, or other organization).
DC. Contributor	Contributor	Person or organization that has had a significant intellectual contribution in creating the document. (eg editor, illustrator, translator, etc.).
DC. Rights	Rights	Reference on copyright, such as URL or link to a good, to a rights management service or a service that will provide information on terms and conditions of access to the electronic resource or DLO.
DC. Give yourself	Date	Date on which the resource was made available to users in its form (electronic) current. This date should not be confused with the *DC* element . *Coverage* associated with a resource only insofar as the intellectual content of the It is related to that date. To determine the date of publication or Publicly available, the W3CDTF format is used (*Web, Data and Time Consortium Format*).

DC. Type	Kind of resource	DLO category in terms of the type of information it represents (eg, image, dataset, text, software, etc.).
DC. Format	Format	DLO data format, used to identify software and possibly also the hardware needed to display it.
DC. Identifier	Identifier	Character sequence used to uniquely identify a resource. For the online resources, these identifiers will be URIs (URL / URN) or other attempts individual (DOI, WEBDAVC, OpenURL, etc.) valid to identify in a way uniform a DLO. For other resources identifier formats can be used such as ISBN, ISSN, etc.

2 Translated and adapted from: DMI. *Dublin core Metadata Initiative* (Revised October 16, 2019). https://www.dublincore.org/ .

1.4.3. Boolean operators

The retrieval of information in a metadata model, reconfigures each term or word to generate relationships with other metadata records and thus retrieve several elements simultaneously. All of this can be accomplished through the use of various models, among them:

- The basic information search system model, which uses a hierarchy theme in which a tree-based reaction is generated, with a central theme axis that unfolds different increasingly specific thematic branches depending on the system needs.

- The complex model based on searching for more specific words or descriptions in documentary records of a database, obtained by the Boolean operators of proximity and time.

On the latter, the Boolean mathematics developed in the 19th century by George Boole defined conditionals of true and false in different mathematical calculations; with the emergence of the bases of data these were applied to information searches [3] , using the semantics in three main connectors (Figure 2) described below:

- **AND** also used as **AND:** defines that, if there are two or more searched words retrieve it in the metadata log, otherwise skip it. It is very useful to dimension a query. Example: Oceans AND Colombia.

- **OR** also used as **O:** this connector defines that, if there is presence of one of the two or more words registered in the metadata, retrieve it, in case there is none omit. It is useful to retrieve large metadata sets to refine with others connectors. Example: OR ports captaincies.

- **NOT** also used as **NO:** this connector is exclusive, defines that if there is presence of the first term and there is no presence of the second retrieve it, in case it does omit it. East it is inverse to the **AND** operator . Example: NOT 2018 oceanography.

[3] In some systems Boolean operators have been graphically included as filter buttons for specific searches or advanced information.

Figure 2. Illustration of the use of Boolean operators [4].

1.4.4. Truncators

Like Boolean operators, these are special character wildcards designed in the 80s of the 20th century with the birth of electronic search systems and the bases of data, to direct the search by a specific pattern; the most common are:

- **Quotation marks “”:** when using these characters, the system will only retrieve records whose texts match the search phrase in the metadata, otherwise it will omit them even if the Words are scattered across the metadata. Example: "merchant marine".

- **Asterisk *** : Search for any record containing words in the metadata content of the same lexical family with difference in one or more characters. Example: For the ocean * search, the system will return records with words such as ocean, oceans or oceanography.

- **Parentheses ():** This connector creates sets with which to segment a search using associative property and set theory. Within each set it is possible to use both boolean operators and truncators. Example: (“merchant marine” AND Colombia) NOT (OR ocean laws *).

[4] Source: The author.

1.5. Final thoughts

Metadata is present in everyday life, in something as simple as a document identity with whose data and through information systems, common points can be generated of millions of these at different times, themes and purposes. So the data and information that are produced can be more useful to the community, if there is a rigorous and standardized record of metadata.

Metadata contributes to solving the need for access to data and information, either by part of an individual or a community; with its elements it is possible to optimize the search and indexing that allow you to reach a document for any reason, such as a phrase exact or single words.

2. EXPERIENCES

2.1. Archeology and recovery of data collected over 40 years aboard Colombian oceanographic cruises

2.1.1. Creation of the Colombian Oceanographic Data Center

Marine research activities on a regular basis in Colombia began to execute in 1969 within the framework of the provisions of Decree 763 of that same year, by means of which The Colombian Oceanography Commission was created (currently the Colombian Ocean Commission, CCO). After four years of field activity at sea, the volume of data was already considerable oceanographic data collected, as well as the number of users that required access to them to different purposes (Dimar, 1973).

With the above, the need to apply the Commission's recommendations became evident Intergovernmental Oceanographic (IOC) of the United Nations Educational Organization, Science and Culture (Unesco) on data management and satisfying the need for access to these (Dimar, 1992). Thus, in 1973 the Special Studies Section of the Oceanography Division of the then General Maritime and Port Directorate (Dimar), formulated the project "Colombian Oceanographic Data Center - Cecoldo" (Dimar, 1973).

The start of the execution of the project in question was planned for September 1, 1973 with a duration of two years (each year a phase), however, Cecoldo had the capacity to provide some of their services before completing the second phase of the project, as evidenced in the report of the Oceanographic Cruise "Ocean III" developed between April and May 1974 in Colombian Caribbean, in which Cecoldo is formally appointed as processor data "by means of a computer program in Fortran IV [...] and a program for stations oceanographic data from the *National Oceanographic Data Center* (NODC) of the United States "(Dimar, 1975), as well as in the successive reports of oceanographic cruises until 1984 (Figure 3) .

At the end of 1978, Cecoldo acquired his own computing capacity by receiving from the Organization of American States (OAS) within the framework of the Multinational Program on Marine Sciences, a WANG minicomputer, and gradually gained independence from the Administrative Department National Statistics Office (DANE), in whose teams the first years of data obtained were processed in Colombian oceanographic cruises (Dimar, 1992).

Figure 3. Cover of the Oceanographic Cruise Report "Ocean IV (Areas 2 and 3) - 1974" published in 1982 [5].

2.1.2. The first experience in archeology and data recovery

Due to the lack of personnel and obsolescence of used computer equipment, in 1984 it was Registered the last formal activity of Cecoldo with the processing of the data of the ERFEN [6] Cruise. IV developed between November 29 and December 12, 1978 in the Colombian Pacific. Seven years later, by order of the Director General Maritime Cecoldo was assigned to the agency organization of the General Secretariat of Dimar (and not of the Systems Section as it had been conceived initially), and it was reactivated on January 31, 1992, beginning the processing and quality control of the data obtained from 23 oceanographic cruises, out of 38 developed to date in Colombia (Dimar, 1992).

The methodology applied in this first retrieval of oceanographic data included gather and document information on the following aspects:

- **Summary of the operation** : generalities of the cruise such as name, ship oceanographic used, time period and description of the area of operation.
- **Sampling** : general description of the variables and measurement or sampling methods.

5 Source: Dimar
6 Regional Study of the El Niño / La Niña Phenomenon.

- **Format** description: detailed description of the variables, measurement units, instruments and / or tables of codes or conventions used.
- **Notes** : detection limits, explanation of the use of alphabetic or numeric characters, problems presented during sampling.
- Station **network:** graphic sample of the sampling station mesh.
- **Graphs and data:** graphical sample of the distribution of observed oceanographic parameters during the cruise, as well as the value of the primary data.

2.1.3. Implementation of "OceanData"

In 1996 the Center for Oceanographic and Hydrographic Research (CIOH) assumed the Center National Oceanographic Data (NODC) before the IOC, and "aware of the need for adequate management to be given to the information produced "(CIOH, 2002) implemented phases I (1996) and II (1997) of the oceanographic and hydrographic information system called "OceanData" (Figure 4). Subsequently in 2001, Phase III was developed within the framework of the IOC International Oceanographic Data and Information Exchange (IODE), "in response to information management and disclosure needs both internally and externally "(CIOH, 2002). It should be noted that no funding was obtained to execute phase IV of the "OceanData" project and with the years this system was relegated.

Figure 4. Appearance of one of the CIOH [7] "OceanData" modules .

The methodology applied during data recovery carried out in the framework of the implementation of "OceanData", was focused on quality control of salinity data and water temperature obtained in developed "Caribe" and "Oceanographic" cruises until 1993 (CIOH, 2002). The tests applied were the following:

[7] Source: CIOH (2002).

- Visual analysis of the data to detect and eliminate, among others, errors in the use of the instrument, such as its descent or rise time.

- Smoothing the data by applying the WFILTER routine of the instrument's SEASOFT software. CTD Sea-bird.

- Selection of samples spatially, applying the BINAVG routine to determine the data at depths meter by meter.

- Filtering data at standard depths (NOAA, 2015) [8] and exporting files to XLS format.

- Upload descriptive information about oceanographic cruises in OceanData.

2.1.4. The "Cenimarc" and "Retrocean" initiatives

In 2004, the then Pacific Pollution Control Center (CCCP [9]) was designated as NODC and started the project "Colombian Marine Information Center, Cenimarc", with the objective of "providing a systematic solution to the problems represented by the redundancy of the information and unstructured processing of data generated by research centers of the DIMAR "(Ortiz and Rico, 2006).

In the framework of this initiative, the CCCP carried out archeology and recovery activities for data from oceanographic cruises developed in the Colombian Pacific, highlighting the following activities (CCCP, 2004):

- Location of data sources in the CCCP facilities.

- Design of templates in XLS format to import the data.

- Development of an application in XML format to load the data into Cenimarc from XLS templates.

- Development of a function to export the results of a query to the *Ocean* format *Data View* (ODV).

[8] They are depths expressed in meters corresponding to the levels defined for the international exchange of measurements. oceanographic in the water column. These are: 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 and 5500 m (NOAA, 2015).
[9] Through article 6 of Decree 5057 of 2009, the name of the CCCP changed to "Center for Oceanographic and Hydrographic Research from Pacific".

- Implementation of the *Marine Environmental Data Information Referral Catalog* (MEDI). metadata management system

For its part, in 2005 the CIOH developed the first phase of the “Retrocean” project, which planned the "Recovery and quality control of oceanographic data", and specifically "information oceanographic of the different cruises developed by the CIOH in more than 30 years of research in the Caribbean Sea, so that it was entered in the Cenimarc database with the respective metadata ”(CIOH, 2005). In Figure 5 the summary of the activities shown archeology and recovery developed at Retrocean.

Figure 5. Methodology used in phase I of the “Retrocean” project [10].

Both the second phase of the Retrocean project and the subsequent phases of Cenimarc did not receive financing, so it was not possible to complete the planned activities.

[10] Source: Adapted from CIOH (2005).

2.1.5. The Atlas of Oceanographic Data of Colombia

In 2013, a specific cooperation agreement was signed between Dimar, Ecopetrol and the Association

Colombian for the Advancement of Science (ACAC), which aimed to "join efforts to carry out the
Oceanographic Atlas of Colombia through the collection, organization, filtering, systematization
and analysis of the oceanographic data collected by Dimar, for consultation through analogous means by the national maritime community "(Convention 5211792, 2013).

As a result, the climatologies of the period 1992-2013 were published in 2015 for the temperature, salinity and density at standard levels of selected depths, from data obtained on board oceanographic ships of the Navy of the Republic of Colombia (ARC) and Dimar in the Caribbean Basin and Pacific Basin of Colombia, and those available in the NODC and the Center Data Center (WDC) of the National Oceanic and Atmospheric Administration (NOAA) in English) from the United States (Andrade *et al.* 2015).

In Figure 6 the workflow followed seen in the data processing to produce the "Atlas of Oceanographic Data of Colombia: 1922-2013 Temperature, Salinity, Density, Geostrophic Speed ", which included archeology and recovery activities such as the inventory and digitization of data and the implementation of a database for an application desktop software. The methodological description and climatology can be found at: http://cecoldodigital.dimar.mil.co/49/.

Figure 6. Flow chart followed in oceanographic data processing [11] .

[11] Source: Andrade *et al* ., (2015).

2.1.6. Cecoldo's strategic positioning

In 2014, archeology and primary data recovery efforts were resumed. by the "Atlas of Oceanographic Data of Colombia", and as an initiative of the Subdirectorate of Dimar Maritime Development, the investment project "Strategic positioning of the Cecoldo "formally executed between 2015 and 2017. This project aimed to" know the quantity and quality of oceanographic data and marine information produced by Dimar from of research and monitoring activities in the Colombian Pacific and Caribbean ”(Dimar, 2018).

Within the specific objectives of the project, it was planned to carry out archeology, recovery and standardization of oceanographic and marine meteorology data produced by Dimar between period 1969 and 2017. The methodology applied to achieve this objective corresponded to the recommended by Hernández *et al* ., (2007) (Figure 7), complementing “PHASE III. Conservation ”with the“ Guide for document transfer and final disposal ”(Dimar, 2016) of the document management process and Dimar content, which includes the guidelines issued by the AGN on the matter.

Figure 7. Phases of the archival methodology for the recovery of historical-oceanographic information proposed by Hernández *et al* ., (2007).

The search and treatment of historical data and information documents (on paper and electronic) was carried out at the facilities of the Oceanographic Research Center and Hidrográficas del Caribe (CIOH) located in Cartagena DT y C. (Figure 8), at the Research Center Oceanographic and Hydrographic of the Pacific (CCCP) located in Tumaco (Figure 9) , and Branch of Maritime Development of Dimar in the city of Bogotá DC, Colombia.

Figure 8. Images of the oceanographic data archeology process at Cioh [12].

Figure 9. Images of the oceanographic data archeology process at Cccp [13] .

[12] Source: Dimar
[13] Idem

Once the data in digital format was obtained and inventoried, the standardization proceeded applying the “Guide to normalization of oceanographic data sets” of Cecoldo (Dimar, 2017), in which apply standards and best practices recommended by the international IODE program of Unesco-IOC, among which the application of ISO standards, formats, vocabularies and the documentation and publication of metadata.

It should be noted that the data quality control carried out in this project focused on guaranteeing that the data sets met the agreed standards, and not the evaluation or validation of each one of the parameters that compose them (that is, the temperature, salinity, oxygen, etc.), therefore, no quality flags were assigned that rate each data obtained as bad, good or questionable. The latter is the subject of initiatives that Cecoldo is currently undertaking.

As can be seen in Table 2, oceanographic (physical and chemical) and

marine meteorology of 118 oceanographic cruises, with the geographic coverage observed in the
Figure 10 whose sampling was carried out by the Navy of the Republic of Colombia or the
Dimar Research Centers. The largest number of cruises have been developed in the basin of the Colombian Pacific and more recently in island areas such as Isla de Malpelo and Isla Gorgona, within the framework of the Regional Study of the El Niño / La Niña (Erfen) Phenomenon, for a total of 63 in a period of 46 years, followed by Caribbean Oceanographic Cruises with a cumulative of 18 in a span 21 years old.

Table 2. Number of cruises whose oceanographic data were recovered in the framework of the "Positioning" project strategic of Cecoldo " [14].

Expedition	Geographic area	Quantity	Years or period
Caribbean Oceanographic Cruise	Caribbean	18	1992 - 2013
CICAR [15] Oceanographic Cruise	peaceful	3	1972 - 1973
Gulf of Urabá Oceanographic Cruise	Caribbean	3	2005 and 2010
Guajira Oceanographic Cruise	Caribbean	2	1990 and 2010
Rosario Islands Oceanographic Cruise	Caribbean	3	1977 - 1978
Ocean Oceanographic Cruise	Caribbean	9	1969 - 1984
Pacific Oceanographic Cruise - Erfen	peaceful	63	1970 - 2016
Petrobras Oceanographic Cruise	Caribbean	4	2007 - 2008
San Andrés and Providencia Oceanographic Cruise	Caribbean	one	2010
Oceanographic Cruise SPOA [16]	Caribbean peaceful	5	2007 - 2008
Oceanographic and Fishing Cruise	Caribbean peaceful	3	1990 and 1996
Seaflower Scientific Expedition	Caribbean	2	2014 and 2016
Colombian Scientific Expedition to Antarctica	Antarctica	2	Austral summer 2014-2015 Austral summer 2016-2017

[14] Source: Dimar-Cecoldo.
[15] Program 'Cooperative Research in the Caribbean and Adjacent Regions'.
[16] Project 'System of Oceanic and Atmospheric Forecasts'.

Figure 10. Geographical coverage of oceanographic and marine meteorology data sets 1969-2017 recovered in the project "Strategic positioning of Cecoldo" [17].

[17] Source: Dimar-Cecoldo

2.1.7. Conclusions

The degree of conservation of paper-based documentation, such as cruise reports, project reports and data sheet folders, provided much of the source primary historical oceanographic cruise data. This is largely due to the document management processes that institutions have been implementing in recent years with the AGN management, and that they have allowed the historical archive to be preserved under conditions specific. Only in a few cases, biological and geochemical data were not recovered since the This component was in charge of other institutions or universities invited to the cruises.

It was observed that, to a large extent, the success of this type of initiative depended on the continuity that was given to the process (that is, the possibility of developing all the objectives or phases planned), of course, supported by financial support and availability of human talent with exclusive dedication to archeology and recovery activities.

A constant was found in each experience and consisted of repeating the process immediately previous; on few occasions it was possible to resume a job developed in advance, probably due to changes in technologies or due to lack of data backup already processed. Despite this, the country currently has a single repository of oceanographic data Colombians, supported by internationally recognized guidelines and protocols in this discipline, which is recognized by the countries that make up the Southeast Pacific region.

For a subsequent initiative of archeology and oceanographic data recovery, it is intended expand the search for information in entities that have participated in the cruises oceanographic, as well as generate links with the products developed from the aforementioned data. The foregoing, so that users can have a global vision of the scope and impact who have had these important scientific expeditions at sea.

2.2. Challenges in archeology and data recovery from the Colombia's largest meteorological monitoring

The General Maritime Directorate (Dimar) began in 2005 the implementation of the Measurement Network of Oceanographic Parameters and Marine Meteorology (RedMpomm), in order to improve the ability to observe and predict climate at different scales, as well as monitoring the meteorological conditions that may affect the country's comprehensive maritime and port security.

The RedMpomm currently consists of 35 meteorological monitoring stations and / or real-time tide gauge, deployed along the Caribbean and Pacific coastline and island areas Colombian, generating on average more than 50,000 daily data. The first seasons were installed in Turbo, Providencia and Coveñas in 2005; in the following years new observation platforms, but it was in 2013 when Dimar made the largest investment with the installation of two tidal stations and eleven meteorological-tidal stations, with the aim of expand the coverage of sea level monitoring, as a key input for the assessment of arrival of tsunami waves to the Colombian coasts by the Dimar Tsunami Warning Center.

More than a decade after starting the measurements, it was evident that the lack of adoption of standards and good practices throughout the data life cycle, that is, planning, acquisition, processing, archiving and publication (CTN Diocean, 2018), had led the network to count with duplicate data, various storage formats, no data in series, and a level of quality that was not accepted by internal and external users.

Therefore, and taking into account that "the main objectives of database management consist of permanently maintaining the integrity of the database and ensuring that it is contains all the data and metadata necessary to respond to the needs for which reason was established, both now and in the future "(OMM, 2011), Dimar appropriated resources and efforts to perform the archeology and data recovery of the RedMpomm. In the following sections describe the activities carried out, which are summarized in the five stages that seen in Figure 11.

Stage 1.	Stage 2.	Stage 3.	Stage 4.	Stage 5.
Search and inventory data	Assessment the series of weather	Standardization of the measurements	Structuring from the base of primary data	Control of quality of data

Figure 11. Stages of the archeology process of meteorological and sea level data of the RedMpomm [18] .

[18] Source: Dimar.

2.2.1. Stage 1. Search and inventory of data

In this stage, the search and inventory of the data found in different equipment was carried out and servers, both from the Dimar headquarters in Bogotá DC, and from its regional offices in Tumaco and Cartagena; Additionally, magnetic and optical devices (eg discs, CD, DVD) in which the preventive maintenance reports of the stations were delivered meteorological and tideographic of the RedMpomm.

The mentioned search results yielded various data sources; these were:

- **Native databases.** This type of database corresponds to a series of directories structured that are managed through the instrument's own software or platform, in this case Hydras 3.
- ***Raw data*** [19] **of the stations** . Refers to native encoded files, unloaded on site each time maintenance was performed at the stations.
- **Exported data** . Corresponds to data in TXT or XLS formats, extracted from the database primary system to be used in specific processes.

2.2.2. Stage 2. Evaluation of the time series

At this stage, the integrity and availability of the databases was verified. native in operation; likewise, the continuity of historical measurements was reviewed comparing recoverable data from the different sources identified.

During this process, different aspects were found that made it difficult to consolidate the time series of the meteorological and tidal variables measured by the RedMpomm; these were:

- **Incomplete primary data.** The *raw data* downloaded on site was limited to periods of approximately three months of data prior to the maintenance date; this because to the configuration of the stations data logger.

- **Missing data.** For various reasons, it was not possible to receive all the data in the series due to the telemetry means that the Network had.

[19] Primary data obtained from the source, in this case, from the memory cards of the observation platforms.

- **Distributed databases.** Given the distributed architecture of the system, local databases for each group of users, which are not synchronized the uncertainty of the data series increased. As a result, there was no unique and consolidated primary database.

- **Data governance** [20] **.** Not having a data governance according to the needs of the Network, the operation and knowledge of the data was not supported in procedures, guides or manuals that show the transformations and uses of these.

Once the data series were consolidated, they were evaluated finding the following errors:

- **Data outside the time range.** Inconsistencies were identified in the start and end dates recorded in some measurements since they did not coincide with the time periods of capture based on the commissioning of the sensors; therefore, they found data from years prior to the installation of the stations and after the deadline of verification (eg data from dates prior to 2005 and data from 2115) (Figure 12) .

Figure 12. Data out of time range in a wind speed series for Buenaventura station [21] .

[20] Data governance “can be understood as a guarantee for the transformation processes' Operational efficiency ',' Knowledge of the data 'and' Use of information '”(MinTIC, 2014).

[21] Source: Dimar-RedMpomm.

Page 33

- **Edited data.** In the absence of data in some periods of time, whether due to damage at the station or due to new data transmission, an attempt was made to estimate the values of the missing data, but no evidence was found of the methodology applied to generate them (Figure 13) .

Figure 13. In red, estimated data in series of stations located in Buenaventura and Cartagena [22].

[22] Source: Dimar-RedMpomm.

- **Impossible data** : periods of time were identified in which data was recorded physically impossible, considering that these are outside the range of sensor measurement (Figure 14) .

(to)

(b)

(c)

Figure 14. Barometric pressure of Buenaventura station: (a) detail of the expected behavior or variability; (b) analysis time period; and (c) detail of impossible data [23].

[23] Source: Dimar-RedMpomm.

2.2.3. Stage 3. Standardization of measurements

Once the data recovery and review was completed, the standardization of measurements like so:

- Deadlines for starting and ending data collection for each station were set, comparing the data found with the historical information available in the worksheets platform life and contractual information.

- The codes of the sensors were standardized, the temporal aggregations were established (that is, how often the measurements are made or frequency of the measurement) and the sensor configuration with the information in Table 3 .

Table 3. Temporal aggregation of the RedMpomm [24] data sensors .

Parameter	Temporary Aggregation
Direction of the wind	10 minutes
Airspeed	10 minutes
Air temperature	60 min
RH	60 min
Barometric pressure	60 min
Global radiation	60 min
Precipitation	10 minutes
Sea level	01 min
Sea temperature	10 minutes

2.2.4. Stage 4. Structuring of the primary database

With the information collected and the decisions made in previous stages, we proceeded to reset the primary database. To do this, a workspace was created in the Hydras 3 software in which the *raw data* found was imported , serving as the basis for the consolidation of the data.

Then, the data available in the system was imported (which previously they were exported under the configurations adopted in stage 3); this activity is accompanied by a visual inspection to identify gaps that could be filled with the data found in other sources, but omitting data calculated, processed or with evidence of modifications or transformations.

[24] Source: Dimar-RedMpomm.

Page 36

As a result of this stage, the primary database was consolidated with about 143 millions of meteorological and sea level data, including discontinuous time series and whose recovery percentages are shown in Table 4 according to the type of platform and place location.

Table 4. RedMpomm meteorological and sea level data recovered as of December 31, 2018 [25].

Platform type	Location	% Total data recovered	Time series annual
Weather Station	Coveñas	69.75%	13
Weather Station	Providence	81.36%	13
Weather Station	Turbo	68.62%	13
Tidal station	Buenaventura	64.80%	10
Weather Station	Buenaventura	82.42%	10
Meteorological and tidal station	Malpelo	79.26%	9
Weather Station	Gorgon	64.20%	8
Meteorological and tidal station	Tumaco	89.09%	8
Weather Station	Barranquilla	84.83%	7
Tidal station	Sailboat port	84.04%	6
Tidal station	Cartagena	80.53%	6
Tidal station	Coveñas	91.43%	6
Tidal station	Providence	97.49%	6
Tidal station	Bolivar Port	59.92%	6
Tidal station	Star Port	74.50%	6
Weather Station	Bolivar Port	67.70%	6
Weather Station	Star Port	78.37%	6
Weather Station	Sailboat port	88.91%	6
Weather Station	Cartagena	90.98%	6
Meteorological and tidal station	Solano Bay	84.45%	6
Meteorological and tidal station	Whales	83.20%	6
Meteorological and tidal station	Naval island	94.55%	6
Meteorological and tidal station	Juanchaco	84.16%	6
Meteorological and tidal station	San Andres	93.82%	6
Meteorological and tidal station	Sapzuro	58.48%	6
Meteorological and tidal station	Serranilla	29.09%	6
Weather Station	Santa Marta	87.72%	4
Meteorological and tidal station	Malaga Bay	84.95%	4
Meteorological and tidal station	Dreamcatcher	18.55%	4
Tidal station	Barranquilla	88.75%	2
Tidal station	Strong Island	89.25%	2
Tidal station	Turbo	97.37%	2
Weather Station	Punta Espada	99.61%	2
Meteorological and tidal station	Puerto Brisa	99.91%	2

[25] Source: Dimar-RedMpomm.

Page 37

National Technical Committee for Coordination of
Ocean Data and Information

2.2.5. Stage 5. Data quality control

Taking into account that "the objective of quality control is to verify whether the value of a data reported is representative of the measurement intended to be carried out and has not been affected for factors not related to it "(OMM, 2011), at this stage we proceeded to control quality of the recovered data. For this, the scope of the qualification of data, for which three types of quality flags (QF) were selected first-level recommended by the international IODE program (Table 5) .

Table 5. Extract of IODE quality flags applied by RedMpomm [26] .

QF code	Meaning	Description
2	Unknown	Not enough information available to determine quality. of the data.
4	Bad	Data has failed in one or more of the control tests documented quality.
9	Missing data	Missing data in the data series.

Next, the recommendations of IOOS (2014) for the control of quality of wind data and those of IOOS (2016) for sea level data, agreeing and documenting four basic quality tests; these are:

- **Test 1. Syntax validation** : It is used to discard the received messages that contain an unknown structure or that the parameter value or sampling date contains invalid characters. As a result, the data is discarded automatically.

- **Test 2. Temporal aggregation** : This test evaluates that the measured data is within of the expected period or time window, that is, according to the temporal aggregation configured for each variable in the system. If they are outside the sampling period they are automatically discarded.

- **Test 3. Missing data** : This test reviews the entire series of data in search of empty fields. Once these fields are identified, they are filled with the value “-99999” and labeled with the quality flag "9".

[26] Source: Dimar-Cecoldo.

Page 38

- **Test 4. Impossible values** : This test applies the recommendation of NOAA, 2015 cited in CTN Diocean (2018), which explains that "in general, quality tests should include the identification of physically improbable values that the variables reach in usual environmental conditions ". As a result of this test, flag "4" is applied. for data outside the sensor's own measurement range (i.e. the value established by the manufacturer), and flag "2" is assigned for the other measurements.

After applying the quality tests in reference to about 16 million data from marine meteorology recovered for the specific variables of relative humidity, temperature of the air, wind speed and direction, and barometric pressure, for all seasons in the periods of available time, the results shown in Figure 15 were obtained.

100%

80%

heh ta 60%

rcen pr P 40%

twenty%

0%

	Humidity Air	Temp. Air	Vel. Wind	Dir. Wind	Pressure
Qf 2	99%	92%	92%	97%	78%
Qf 4	one%	8%	8%	3%	22%

Figure 15. Result of the RedMpomm marine meteorology data qualification process 2005-2018 [27].

The largest amount of data rated QF "4" or "bad" data corresponded to the parameter atmospheric pressure, probably associated with the effect of the conditions of the sampling area on the measurement sensor; For all the other parameters, data classified as QF "2" were obtained by above 90%, that is, data that will have to undergo additional testing and analysis to determine its quality, in application of the WMO recommendation (2017) "Quality control guarantees the quality and consistency of the data obtained, which is achieved through a set of painstakingly designed procedures that focus on good maintenance practices, repair, calibration, and data quality checks. "

27 Source: Dimar-RedMpomm.

Page 39

2.2.6. Final thoughts

Valeh Nazemoff, Senior Vice President and Co-Founder of Acolyst explains that "the data is the An organization's most important asset, and without standards and data quality, the organization it works " [28] ; For this reason, it is important to develop initiatives to increase the value of data. which is achieved by documenting, formalizing and applying procedures, guides and good practices in the entire life cycle of the data, from its planning to its storage, conservation and provision; in this way, data governance will not be subject to the discretion of the administrator of turn.

It is important to take into account during the design of any information system that includes ocean-atmospheric monitoring variables, the implementation of a "primary" database that contain the data as-is, without validations or corrections, and with no adjustments other than those of its format transformation for storage; the above, to guarantee the data traceability. Likewise, controls (physical and logical) must be established to maintain the integrity of said database.

The Colombian Oceanographic Data Center [29] contributes to meeting the objectives of the Unesco-IOC IODE program at the national level, among them is the management of practices for the management, exchange and access to data; therefore it is advisable to establish communication with Cecoldo during the planning of archeology and recovery initiatives type of data.

The Dimar RedMpomm plans to develop archeology and retrieval of oceanographic data for the wave parameter. It is estimated to recover a series of up to 10 years of variables such as height, period, direction and spectral dispersion of the waves, to seven monitoring points located in Tumaco, Isla Gorgona, Buenaventura, Bahía Solano, Turbo, Barranquilla and Puerto Bolívar.

[28] Recovered from https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/bid/406201/qu-es-el-gobierno-de-datos-y-por-qu-lo-i need .

[29] Available at: https://cecoldo.dimar.mil.co/ .

2.3. Preservation process of the aerial photography collection history of Colombia: inventory, digitization and disposal for consultation and development of applied studies

Carantón and Álvarez (1998) defined an aerial photography or aerial photography as the “image of a terrain captured from an airplane or any other aircraft equipped with photographic cameras special for this purpose. It is the complete and detailed record, at the moment of taking, of the elements that make up the geographical environment, that is, of the physical phenomena of the natural environment and of the complex of actions that man carries out in this environment ”. Colombia is one of the first countries of the continent in using aerial photographs for the cartographic survey of its territory, IGAC being the leader of this task.

Since its inception, the Agustín Codazzi Geographical Institute (IGAC) was conceived as an entity cartography producer, where its main inputs over time have been the aerial photographs. It all started in 1902 when a working group called in that moment "Office of Lengths and Borders" attached to the Ministry of Foreign Affairs, which only in 1910 he began to act, and twelve years later he began photogrammetry work working in together with the aviation company Sociedad Colombo Alemana de Transporte Aéreo (SCADTA), where Approximately 700 aerial photographs were captured on the Magdalena River. In 1935 appoints as Military Geographical Institute, and fifteen years later adopts the name that has currently in memory of the Italian cartographer Agustín Codazzi (IGAC, 2014).

This process has been going on for more than 84 years; at first they were captured images of the Colombian territory with the use of analog cameras, obtaining negative films, which were stored by means of rolls. In 1956 the construction of the infrastructure that today today it is known, and in 2007, 51 years later, the IGAC through the project "Systems improvement mapping of the Colombian territory ”and with funding from resources donated by the Community European, acquired a digital aerial camera system for taking digital aerial photographs, in addition of the acquisition of three photogrammetric scanners, with the aim of preserving the historical heritage accumulated from rolls of aerial film negatives (IGAC, 2014).

The historical collection of aerial film negative rolls, formed from 1935 to 2007, is made up of 5,734 rolls, each of which contains approximately between 200 and 270 photographs on plastic media. In total the series of images brings together approximately 1,400,000 monochromatic negative (black and white) aerial photographs.

The plastic support of the negatives corresponds mainly to cellulose acetate, Ethylene polyterephthalate (PET) and resin paper; and to a lesser extent cellulose nitrate. The image of the photographs is made of silver halide bonded with gelatin.

It is important to highlight that this type of photography has great strength and great weakness. The strength lies in the chemical and physical stability of the imaging technique, since the halides of silver bonded with gelatin are very stable over time. If the image support is paper or resin paper, as in this case, photography is even more stable, due to the chemical structure of the paper.

Weakness arises when the support is made of plastic, either cellulose nitrate, cellulose or PET; in this case plastics undergo autocatalytic processes, emanating some of the substances during aging processes; these processes cannot be stopped and if the emanation levels are quite high, the process is noticeably accelerated. For this reason it is very It is important to ensure adequate storage conditions for this entire collection, a condition which greatly slows down the process, but allows the preservation of the photographic material by a long time.

Another important factor considered in the process of preserving information aerial photographs, is the technological risk related on the one hand, with the limited availability of photogrammetric scanners to digitize the images of the rolls of photography negatives aerial in the IGAC, and on the other, the fact that these teams are no longer being produced, which it makes it increasingly difficult to achieve it in the market and makes maintenance difficult during its useful life. This risk implies a significant challenge for IGAC in terms of digitizing the missing aerial photography negatives in the shortest possible time and thus guarantee the preservation of the information collected that is part of the institutional heritage.

It is worth mentioning that to safeguard the information of historical aerial photographs of the territory national, it is required, on the one hand, to digitize the material, and on the other, to guarantee its conservation in the original format in which they were taken, that is, on the rolls of aerial film negatives, arranging the adequate conditions for its storage, advancing the procedures technical and administrative so that this file is declared by the AGN as a property of cultural interest.

The process of preparation, handling, inventory and storage of the material that will be digitized, is carried out under the guidelines and technical advice of the AGN; basically consists of stages described below.

2.3.1. Preparation and storage

The material is currently located in an IGAC vault, built 60 years ago. approximately. The entrance to the vault is made through a metal security door that allows greater control over people's access, consistent with the fact that it is information a delicacy of high national importance that lies there (Figure 16) .

Figure 16. Photo of the poster of the entrance of the warehouse of rolls [30].

The vault is a two-story enclosure comprised of a planoteca on the first level (Figure 17) and in the second the oldest films in the collection are stored. In an organized way, in The packages that are tubular closed boxes with the rolls of air film negatives, which are handled for inventory or digitization (Figure 18).

Figure 17. Internal photo of the winery (Planoteca) [31].

[30] Source: Colombian Spatial Data Infrastructure (ICDE).

[31] Source: IGAC Geography and Cartography Branch.

(to) (b)

Figure 18. (a) Internal photo of the warehouse with the shelves of rolls and (b) photographic roll [32] .

Modifications have been made in this venue to meet the recommendations of curators experts in this type of preservation regarding the storage and handling of analog images, according to the AGN guidelines; for example, one of these recommendations is related to the enabling and maintenance of motors that absorb and extract air, seeking constant flow inside the storage vault; Also within the recommendations are the oriented so that the shelves are of plastic and non-metallic material, and the paintings cannot be of oil; all depending on the preservation of the photographic rolls.

A new inventory is currently being generated under the parameters established by the AGN, for which the IGAC has arranged the required personnel. Likewise, the adaptation of an area for the handling of rolls, where several extractors and some charcoal repositories which, together with daily spraying of a calcium chloride solution dissolved in distilled water, they help the decantation of the particulate materials emanating from the rolls when removing them from their container containers.

Another important aspect corresponds to the mandatory clothing for handling, in this case consists of glasses, special masks, gloves, and white coats for officials who they interact in these spaces.

[32] Source: IGAC Geography and Cartography Branch.

In the storage vault, located on the second floor of the IGAC main building, a conveyor belt (Figure 19) is installed through which the rolls of mechanically to the lower floor, where the area enabled for handling the rolls is located and its inventory in the old photographic laboratory. In this space are the scanners photogrammetric for digitalization of aerial photographs; also in this area, in several of the old dark rooms, there are still photographic equipment, which due to its approach to Technical obsolescence will be arranged in the future in a cartographic museum, planned to conform in said area.

Figure 19. Roll conveyor belt in warehouse [33].

2.3.2. Handling

In order to inventory or digitize the rolls, they are transported through the belt conveyor the day before the one to be operated, placing them on a covered table with a special cloth; the rolls are kept uncovered all night to expel gases accumulated. If they become very old scrolls, they are left from the Friday before the week in which they will be manipulated, with the aim that during the weekend they are ventilated and the gases.

Additionally, there are constantly turned on fans for air circulation; precisely one of the greatest agents against preservation is dust, which is It is important to manage an environment as sterile as possible. At various corners of the work area, they place repositories with carbon, to help the decantation of the particulate materials, which Helps mitigate odors, reducing the health risk of officials who perform processes associated with inventory or digitization.

[33] Source: ICDE

2.3.3. Inventory

An official is responsible for verifying the status of each roll and collecting the image data such as flight, date, number of scenes per roll, among others (Figure 20); this person permanently use gas mask with glasses, since the chemicals that expel the rolls are powerful. For the inspection of the rolls a special machine arranged for this purpose with glass is used through which LED type light is projected, it is also equipped with transport handles for the film for inspection of each photograph per roll. Simultaneously, a digital form where the result of the inspection carried out is recorded.

Figure 20. Roll handling [34].

2.3.4. Digitization

The digitization process is carried out with a photogrammetric scanner; depending on its type you can scanning done manually or automatically. In the case of the automatic scanner, the equipment uses on permanently, because scanning can take up to three days per roll, running 24 hours. After this and having the images digitized, using a software program is organized and classified.

Digitization is carried out from two work fronts:

- **Work plan** . Refers to the digitization plan for full rolls of negatives from films. In this process, automatic scanners are used (Figure 21) and a Control record of the aerial photographs arranged in digital format.

[34] Source: IGAC Geography and Cartography Branch.

Figure 21. Digitalization room with automatic scanners [35] .

- **On demand.** Corresponds to the attention of specific user requirements that request a limited number of images from the query of the aerial photographic archive.

For this work a manual scanner is used.

It should be noted that the images are mostly available without georeferencing, since this It is one of the activities that IGAC will carry out as part of the preservation process of the information.

2.3.5. Documentation

Currently metadata of aerial photography images are manually documented under a profile of the ISO 19115 standard, using the catalog software that IGAC has for such an end.

2.3.6. Disposal and use

After digitized, the images are stored and arranged in the National Bank of Images (BNI [36]), a solution developed by IGAC to allow online consultation of images of remote sensors, as well as other derived products that the Institute has.

35 Source: ICDE.
36 Available at: http://www.bni.gov.co/

2.3.7. Final thoughts

To date, about 331,800 images corresponding to an advance have been digitized approximately 23.7% of more than 1,400,000 available, which as mentioned, are contained in all the rolls stored in the IGAC. Taking into account that the treatment of each aerial photography takes about 23 minutes and the care that must be taken with the handling of the material and especially the older rolls, it is estimated that the task of digitizing the remaining rolls it will take several more years. Considering the time and investment required, IGAC is managing Inter-institutional resources and alliances for other entities to help preserve the aerial photographic archive of the nation, in search of joining efforts to arrange this important information for different uses.

The importance of these inputs for the nation is highlighted, being the only archive of this nature in the country, made up of invaluable information for numerous studies and a reference history of the Colombian territory. It can also be considered the starting point to get to know the geographical dynamics that the country has had, from the point of view of the multitemporal change in the use and occupation of the territory.

Finally, the foundation of the preservation and conservation process must be understood in its true scope, it's not just as a digitization process to safeguard information aerial photography recorded on film reels in digital medium, but also to preserve this archive as a historical and cultural asset of the nation, reducing its deterioration and guaranteeing its adequate preservation for future generations.

2.4. Review of archeology and information retrieval for the "Petroleum Information Bank"

Conducting archeology and information retrieval for the Information Bank Petrolera (BIP) has been of great importance; Since its inception, methodologies based on techniques that refer to the systematic process of recovering data recorded in media obsolete, likewise, the information has been secured by storing it in computer systems that day by day they become more technologically sophisticated.

As diverse as geoscientific information is, have been the processes and actors that have taken to the BIP to its current state in the Colombian Geological Service. In this effort they have been present professional experts in the production, management, organization and use of data such as doctors, geoscientists, among others, and topics such as search, recognition, recovery, completeness and preservation of information. Then it presents an overview of these processes.

2.4.1. The first "Database"

In July 1983, the idea of organizing the first “Data Bank” was born at Ecopetrol, and for this, the search and classification of media that were in different places like the same began Ecopetrol, where its officials had seismic or well information under their responsibility recorded in devices such as tapes 21, 9 and 7 *tracks* or tracks, cartridges, among others (Figure 22) ; It was also used to search for means that could be in operating companies, companies of services, seismic data processing centers, etc.

Figure 22. Type of media from which data was retrieved for BIP [37].

[37] Source: SGC.

With this exercise, it was possible to gather approximately 65,000 magnetic media with data seismic (primary and processed). Initially, these were stored in the city of Bogotá. where an entire management process based on archeology and data recovery began for that the information could be made available to the entire interested community.

Although there was already talk of a "Data Bank" since 1983, it was until 1985 that the story began as “Petroleum Information Bank - BIP”, when Ecopetrol undertook the task of building a adequate space to keep all recovered media in its best condition, and decided to do so in the Mansilla district of the Municipality of Facatativá, Cundinamarca, Colombia, with the Cintoteca [38] NRP "Nelson Rodríguez Pinilla" as it is known today (Figure 23) .

Figure 23. Overview of the Cintoteca located in the department of Cundinamarca [39].

Once the construction of the Cintoteca was finished, it had safe and sufficient spaces so that in 1987 the 65,000 magnetic means, as well as similar means, moved from Bogotá which especially represented the observer's reports, elevations, coordinates, maps and other documents, the result of the seismic surveys that had been carried out up to that moment in the country. Once this was done, in 1990 the data archeology was continued obtaining information from the well logs complementary to the one that was already in the seismic Belt, maps and documents technicians. In this way, it was possible to have all the E&P information in one place.

2.4.2. Database development

In 1998, with all the information under control at the Cintoteca, the need to take to another state, the media and similar information that was stored there. Then it was undertaken a new initiative for archeology and data recovery aimed at the system in place nowadays. From then until 2002, practices were applied that allowed preserving the magnetic and analog media to bring all information to modern recording media data such as *Digital Line Tape* (DLT) and *Linear Tape-Open* (LTO) and in digital format (Figure 24) .

[38] The Cintoteca NRP preserves all the physical means of technical information on hydrocarbon exploration and production. It is there where the process of delivery of said media by the EPIS (*Exploration and Production Information Service*) for its final disposition.

[39] Source: SGC.

Figure 24. Transformation of data storage media and technical documents E&P [40].

Programs were developed that allowed the recovery of data that was, even, in formats unknown that did not correspond to those handled by the *Society of Exploration Geophysicists* (SEG). The means that at the time presented fungi, became obsolete, and for their treatment they applied baking techniques in specialized ovens, cleaning with equipment that removed the impurities (without damaging the tapes), and preservation based on scanning with Optical Recognition of Characters (OCR), which ultimately enabled the development of a database from where you could have access to all the information available in digital format.

In 2003 the national government decided to create the National Hydrocarbons Agency (ANH) and then observed the need to transfer to this entity the BIP that had matured under the Ecopetrol administration; constituting this in the first transfer of the BIP. In 2011 the The national government evaluated making a new transfer from the information bank, this time to the SGC, for which it established an execution period of five years, thus developing the BIP operation in ANH until November 2, 2016.

2.4.3. The BIP today

Due to the transfer of the BIP from the ANH to the SGC, it was necessary to resort again to archeology and data recovery techniques applied to difficult to read media, either by obsolescence of storage devices or due to the state of the data. Is necessity was born from the identification of BIP information required by users, which, Although it was already in digital repositories, downloading it in its original format was damaged or encrypted. For these cases, a process of searching for old media was started, which was they store in the NRP Cintoteca.

40 Source: SGC.

Once the type of media has been identified, specialized companies that have with older media readers, and with the appropriate drivers with data recovery rates and / or files above 80%. Following this strategy, suppliers with equipment were found available for media read data seismic and well recorded, such as tapes of 9 *tracks* , 3480, 3480E, 3590, 3592 cartridges, 8 and 4 millimeter exabytes, among others.

Currently, the SGC develops archeology and data recovery to the extent that these are required; the Cintoteca NRP is one of the assets of the BIP that stores more than 1.3 million physical means, from which the requested information can be retrieved, thus being able to attend user demand.

The goal of the BIP is to ensure that all the information that the information bank has is most complete, reliable, complete and of the best possible quality. All of the above, taking into account that technology projects focused on achieving recovery are currently being developed and 100% completeness of the information so that it is available online.

3. PERSPECTIVES

3.1. Archeology and recovery of park biological records National Naturals of Colombia

3.1.1. General

National Natural Parks of Colombia (PNN) developed and put into operation the monitoring information of "National Natural Parks of Colombia SULA" since 2013, being This is an online computing tool [41] , under the architecture of the Google App Engine, which supports the process of structuring and storing monitoring data, as well as the visualization of status, pressure and response indicators of the National Natural Parks of Colombia (PNN, 2018).

The system was created taking into account monitoring tools, however since 2018 The subject of scientific research was added, which created the need to advance a process of archeology and retrieval of historical information on point or baseline studies, including habitat of terrestrial and coastal marine ecosystems, originating from characterizations, reports from contractors / consultancies, research permits or guarantees, scientific articles, projects, prevention, control and surveillance tours. All of the above in order to integrate and make such information available for PNN.

The information in SULA is organized in thematic folders, data structures and batches of data (Figure 25) . In the case of scientific research, the following are integrated into this topic data structures that are based on the DarwinCore standard provided by the Information System of Biodiversity of Colombia (SiB Colombia, 2019), and whose templates were adapted to the needs organization of PNN research information, as follows:

- **Species lists.** They are compilations of species or taxa within a data set It may have a thematic or regional scope. It is used to make the species lists recorded in both primary and secondary information. They are organized by groups biological or protected area and geographic coordinates are not included.

- **Biological records** (observations and / or occurrences). They correspond to direct sightings in the field of any specimen or to temporary collections where the individual is released in the same place it was found. Includes taxon and geographic coordinates.

41 Available at: http://sula.parquesnacionales.gov.co/

Figure 25. Information folders in SULA [42].

3.1.2. Progress achieved

Currently in the subject of research 16 protected areas have linked information from 40 batches of data and 23,942 records for the biological groups shown in Table 6 . New information is constantly being entered using as additional strategies to work from area teams, support from volunteer rangers and interns.

Table 6. Biological groups integrated in the research theme of SULA [43].

Protected areas structure		Biological group	Data batches	Records
List of species	Serranía de los	Fauna and Flora	2	1821
	Yariguies	Fish, mammals and plants	3	305
	Tinigua Central	Birds	one	191
	Churumbelos	Mammals	one	3505
	La Paya	Fish, birds, flora	6	852
	Shit	Herpets	one	32
Records biological	The cocuy	Stenocercus lache, Espeletia curialensis	2	346
	Estoraques	Entomofauna and plants	2	6657
	Cliffs	Herpets	one	55
	Guacharos	Birds	one	48

42 Source: PNN

[43] Source: PNN

Page 54

Protected areas structure	Biological group	Data batches	Records
Guanenta	Polylepis quadrujuga, Espeletia Congestiflora, Espeletia boyacensis, Inca spelled	7	8306
The Flamingos	Birds	one	86
Picachos	Fishes	one	25
Sanquianga	Fishes	one	24
Jungle of florence	Herpets	one	165
Sumapaz			
Tatama	Frailejones baseline	2	779
	Coleopters, Dragonflies, herpets, orchids	4	737
Tinigua	primates	one	8

The task is great and corresponds to a first contribution in the work of systematizing knowledge historical base of protected areas of the national order. It is expected to be able to advance in strengthening of the documentation centers of the protected areas and the territorial directorates, beginning with an inventory of both documentary and database information, in order to have a clear panorama of the current state and to be able to advance in the management of knowledge of the history of National Natural Parks of Colombia.

3.2. Archeology and information retrieval for the largest inventory geoscientist of Colombia

3.2.1. General

The information collected in the form of products by the Colombian Geological Service (SGC) is the fundamental food to nourish the geoscientific and nuclear information services of the entity, which day by day provides information from specific requests, which seek to resolve problems in communities, industry and government at the national and territorial levels. No However, the information produced does not fulfill its purpose if it is not provided for consultation and if it is not widely disclosed.

That is why on August 28, 2017 the SGC started the operation of the Integration Engine of Geoscientific Information (MIIG) [44] , conceived in the Enterprise Architecture to facilitate and support the missionary processes of generation and supply of information to users from different sectors such as they are the scientific, academic, public, private, and general citizenship. The MIIG allows cataloging and organize the information in centralized repositories that facilitate their access to users through of search, query and download mechanisms for the largest geoscientific inventory in Colombia, Composed of missionary studies, downloadable and editable files produced over 100 years information production.

The MIIG stores approximately 64,736 metadata records corresponding to 6.51 TB of files of the following types: (i) 10,679 study packages made up of 21,716 documents between reports, memoirs, documentary annexes, field books, figures, etc .; and (ii) 32,341 maps (cartography, plates, atlases, etc.).

Grouped by historical periods, the study packages available in the MIIG are:

- 7997 produced between 1916 and 1999, containing 14,080 documents and 22,956 maps in Pdf format.
- 2,682 produced between 2000 and 2019, containing 7,636 documents in PDF format and editable files with .DOC and .XLS extensions, as well as 9,385 maps also in format PDF and native or editable files in extensions such as .SHP, .COB, .E00, .MXD, .GDB, among others, according to the evolution of GIS formats.

As an example of the historical information available in MIIG, Figure 26 shows the results of a search carried out on the platform, referring to documents with information on the “Napipí Interoceanic Canal” project generated in the 1920s.

[44] Available at: https://miig.sgc.gov.co

Figure 26. Aspect of the MIIG platform to search for geoscientific information in the country [45].

3.2.2. Progress and goals

Be attentive to the archeology and data recovery processes in the face of changes technological and information preservation needs, it is an arduous work that avoids the loss of these important assets to the nation. It involves the use of advanced technologies with parameters well defined collection, support and security, which are achieved with the participation of a team of professionals committed to these activities.

[45] Source: SGC

For this reason, since 2018 the SGC has been developing an archeology initiative and information retrieval, which includes improving metadata content and data and information existing in the MIIG inventory, taking into account appropriate parameters of quality, completeness and visualization of documents, attachments and maps that users access through of this tool.

Archeology and recovery activities have focused on:

- Detailed review of documentary and cartographic elements listed in tables of contents of the main documents.

- Search in different documentary funds in physical support, such as documents originals, map library, library, in which the historical missionary memory rests institutional.

- Digitization of information generated between 1916 and 1999, using OCR technology and PDF format (Figure 27) , which allows you to retrieve content through the search engine of the MIIG platform.

Figure 27. Document produced in 1929 scanned with OCR functionality for easy retrieval and attached map Available in PDF format [46] .

[46] Source: SGC.

- Optimization of the information organization process and its respective files native or editable.

- Cataloging of information in areas of knowledge such as basic geosciences, mineral resources, geological hazards, hydrocarbons, nuclear issues and laboratories, as well as in lines of research in geology, geomorphology, hydrogeology, geothermal energy, metallic, non-metallic and energy minerals, hydrocarbon resources, among others. The above in order to facilitate the search for information from terms or words keys, applying boolean operators, and by interacting with a map - viewer geographic (search for resources that in their attribute values totally or partially with the given coordinates).

As a result of the implementation of the mentioned activities, there is currently a 22% advance with respect to the total inventory of records of documentary and cartographic information, equivalent to 64,736 files reviewed and updated. It is planned to cover 50% of the information by the end of 2020 and by 2022 have completed the process of reviewing the quality of QMS data, metadata and geoscientific information.

4. BIBLIOGRAPHY

Andrade, CA, Rangel, OE and Herrera, E. (2015). *Atlas of the Oceanographic Data of Colombia 1922-2013 Temperature, Salinity, Density, Geostrophic Speed* . Bogotá, Colombia: Address General Marítima-Ecopetrol SA Ed. Dimar. Recovered from http://cecoldodigital.dimar.mil.co/49 .

Carantón, D. and Álvarez, VJ (1998). *Application of aerial photographs in geography* . Bogota, Colombia: Agustín Codazzi Geographical Institute. http://biblioteca.igac.gov.co/janium/Documentos/1-00305.pdf.

CCCP. (2004). *Project "Colombian Marine Information Center, CENIMARC".* Tumaco, Colombia: General Maritime Directorate.

CIOH. (2002). *Project "Expansion of CIOH Data Center services" Phase III / IV* . Cartagena, Colombia: General Maritime Directorate.

CIOH. (2005). *Project "Recovery and quality control of oceanographic data CIOH - Back off "Phase I / II.* Cartagena, Colombia: General Maritime Directorate.

Cooperation Agreement 5211792 for the development of specific and technological activities between the Ministry of National Defense - General Maritime Directorate, Ecopetrol SA and ACAC. Bogota DC, Colombia. July 26, 2013.

CTN Diocean. (2018). *2018 Reference Manual on Best Data Management Practices Oceanic* . Bogota, Colombia: Address general Maritime. http://cecoldodigital.dimar.mil.co/2317/1/cco_2539-2212_manual_referencia_mpgdo_3_2018.pdf .

Dimar. (1973). *Project "Colombian Oceanographic Data Center"* . Bogota Colombia: General Maritime and Port Directorate.

Dimar. (1974). *Ocean III Oceanographic Cruise Report - 1974* . Bogotá, Colombia: Address Maritime and Port General.

Dimar. (1975). *Ocean IV Oceanographic Cruise Report Areas 2 and 3 - 1974* . Bogota Colombia: General Maritime and Port Directorate.

Dimar. (1992). *Project "Reactivation of the Oceanographic Data Center of Colombia"* . Bogota, Colombia: General Maritime Directorate.

Dimar. (2016). *G1-GUI-001 " Guide for transfers and final disposition of documents", version 0* . Bogotá, Colombia: General Maritime Directorate.

Dimar. (2017). *M8-GUI-001 " Oceanographic Data Set Standardization Guide", version 4* . Bogotá, Colombia: General Maritime Directorate.

Dimar. (2018). *Annual Report of the Colombian Center for Oceanographic Data (Cecoldo) 2017* . Bogotá, Colombia: General Maritime Directorate. http://cecoldodigital.dimar.mil.co/2184/

DMI. *Dublin core Metadata Initiative* . (Revised October 16, 2019).

https://www.dublincore.org/

Eco, H. (1984). *The Name of the Rose* . Barcelona: Lumen.

Hernández Jaimes, JL; Ortiz Martínez, RV and Suárez Pinzón, I. (2007). *Archival methodology for the recovery of oceanographic information from the Colombian Pacific* . CCCP Scientific Bulletin (14). pp. 123-150.https://doi.org/10.26640/01213423.14.123_150

IGAC. (2014). *IGAC 80 years.* (Revised November 25, 2019). https://www.igac.gov.co/es/contener/igac-80-anos.

IOC. 1999. *International Global Oceanographic Data Archeology 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). *Standard metadata and information retrieval, problems and applicability in digital libraries* . Barcelona, Trea Editions.

Méndez Rodríguez, E. (2003). *Metadata and information retrieval: standards, problems and applicability in digital libraries. Documentation Annals, 295* . https: // doi-org.ezproxy.javeriana.edu.co/10.6018/analesdoc.6.0.3591

MinTIC. (2014). *G.INF.06 Technical Guide - Data governance. Version 1.0 of December 30, 2014* . Colombia: Ministry of Information Technology and Communications. https://www.mintic.gov.co/arquitecturati/630/articles-9258_recurso_pdf.pdf .

NISO. (2019). *Implementing Linked Library Data* . (Revised October 10, 2019). https://www.niso.org/events/2019/11/implementing-linked-library-data .

NOAA. (2015). *Contents of World Ocean Atlas 2001 CD-ROMS* (Revised October 15, October 2019). https://www.nodc.noaa.gov/OC5/WOA01/cdwoa01.html.

OMM. (2011). *WMO Climatic Practices Guide-N ° 100. Edition 2011* . Geneva, Switzerland: Organization Meteorological World. http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf .

OMM. (2017). *Guide to Meteorological Instruments and Methods of Observation WMO-N ° 8. Edition*

2017 . Geneva, Switzerland: Organization Meteorological World. https://library.wmo.int/doc_num.php?explnum_id=3664.

Ortiz Martínez, R. and Rico Lugo, H. (2006). *Software Tool for the Administration of the Colombian Marine and Coastal Information* . CCCP Scientific Bulletin (13). pp. 3-16. ISSN 0121-3423. https://doi.org/10.26640/01213423.13.3_16.

PNN. (2018). *Instructions for the administration and management of monitoring information* . 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. ACRONYMS

ACAC	Colombian Association for the Advancement of Science
AGN	General Archive of the Nation
ANH	National Hydrocarbon Agency
ARC	Navy of the Republic of Colombia
BEEP	Oil Information Bank
BNI	National Image Bank
CECOLDO	Colombian Oceanographic Data Center
CENIMARC	Colombian Marine Information Center
BCC	Colombian Ocean Commission

CCCP	Pacific Pollution Control Center
Cccp	Pacific Oceanographic and Hydrographic Research Center
CICAR	Caribbean Cooperative Research Program and Adjacent Regions.
CIMI	*Consortium for the Interchange of Museum Information*
CIOH	Oceanographic and Hydrographic Research Center
Cioh	Caribbean Oceanographic and Hydrographic Research Center
IOC	Intergovernmental Oceanographic Commission
CPC	Colombian Pacific Basin
CTD	*Conductivity, Temperature, and Depth*
CTN Diocean	National Technical Committee for Ocean Data and Information Coordination
DANE	National Administrative Department of Statistics
DC	*Dublin Core*
DCIM	*Dublin Core Metadata Initiative*
DIMAR	General Maritime Directorate
DLO	*Document Library Object*
DLT	*Digital Line Tape*
DOI	*Digital Object Identifier*
EAD	*Encoded Archival Description*
EPIS	*Exploration and Production Information Service*
ERFEN	Regional Study of the El Niño / La Niña Phenomenon
E&P	Exploration and Production
GILS	*Government Information Locator Service*
GODAR	*Global Oceanographic Data Archeology and Rescue*
ICDE	Colombian Spatial Data Infrastructure

Page 63

IGAC	Agustín Codazzi Geographical Institute
Invemar	Institute of Marine and Coastal Research
IODE	*International Oceanographic Data and Information Exchange*
IOOS	*Integrated Ocean Observing System*
ISBD	International Standard Bibliographic Description
ISBN	*International Standard Book Number*
ISSN	*International Standard Serial Number*
LTO	*Linear Tape-Open*
MARC	*Machine Readable Cataloging*
MEDI	*Marine Environmental Data Information Referral Catalog*
MIIG	Geoscientific Information Integration Engine
NOAA	*National Oceanic and Atmospheric Administration*
NODC	*National Oceanographic Data Center*
OCLC	*Online Computer Library Center*
OCR	*Optical Character Recognition*

ODV	*Ocean Data View*
OAS	Organization of American States
OMM	World Meteorological Organization
PET	Ethylene polyterephthalate
PNN	National Natural Parks of Colombia
QF	*Quality Flag*
RedMpomm	Network for Measurement of Oceanographic Parameters and Marine Meteorology
SCADTA	German Colombian Air Transport Society
SEG	*Society of Exploration Geophysicists*
SGC	Colombian Geological Service
SiB	Biodiversity Information System
SPOA	Oceanic and Atmospheric Forecast System
URI	*Uniform Resource Identifier*
Unesco	United Nations Educational, Scientific and Cultural Organization
W3C	*World Wide Web Consortium*
W3CDTF	*Web, Data and Time Format Consortium*