

# Meta-Database of Marine Research in Mexico: Trends and Applications

Draft

*Palacios-Abrantes, J; Arreguín-Sánchez, F; Cisneros, A; Cisneros-Mata, M.A; Rodríguez, L.*

*4/12/2017*

## Introduction

Research and management of marine resources increasingly depends on various biological, ecological, social, and economic data (Mokrane & Parsons, 2004; Reichman et al., 2011). The availability of data is often perceived as a gap in advancing research and policy discussion [cita]. However, in many cases, this is largely a result of the lack of knowledge about the availability of these data (Reichman et al., 2011). In Mexico, numerous information covering the seas and coasts can be found in academic institutions, government, and NGOs located (physically) both inside and outside the country. Scientific institution projects like DataMares from the University of San Diego (UCSD) and UNIMAR from the National University of Mexico (UNAM) have long shared data through their portals. Non governmental agencies programs like *Monitoreo Noroeste* provide metadata of marine saplings. Governmental institutions like the National Commission for Biodiversity (CONABIO) hosts one of the largest biodiversity databases of the country. Recently the Mexican president released a decree by which he established a policy on Open Data DOF: 20/02/2015. The initiative Open data (*Datos abiertos*) provides a legal framework for any Mexican citizen to access public data generated by the Mexican government (<https://datos.gob.mx/>). While these are just some of the data sources available for marine data in Mexico, it provides an idea of the variability that could exists in terms of data structure (e.g units, scales, formats). Moreover, in many cases, data lacks basic information compromising its utilization. Having publicly accessible description on existing data is a huge step towards increasing collaboration and innovative research (Reichman et al., 2011; Heidorn, 2008).

Data availability is key not only to better understand Mexico's marine and coastal environments, but to identify knowledge gaps so that research can be prioritized (Cisneros-Montemayor et al., 2016). This will facilitate furnishing management and conservation policies, for example, for marine habitats and fisheries resources vulnerable to climate change. Building tools for effectively managing collected data and access while respecting researchers rights to publication is an important step towards data preservation and stewardship (Vincent, et al., 2010). The creation of a meta-database of marine research in Mexico will foster collaboration and eases the process of informing best policies relevant to any community or region in the country.

A meta-database is a documentation of the sources of information instead of a database of the actual data. It describes the “who, what, when, where and how” of the data collected (Michener, 2006). Having a meta-database helps understanding the data, increases data longevity, and enhances collaboration among scientists (Michener, 2006). Moreover, if well documented, the metadata itself could be used to explore science and policy questions (Cisneros-Montemayor et al., 2016; Michener, 2006).

The current project aimed to create a meta-database for oceanographic, ecological, economic, fisheries and social data for marine ecosystems and marine-related sectors of Mexico. It subsequently analyzed the information collected to determine trends and data gaps in Mexico. The final meta-database is an open source repository for public consultation and sharing.

# Materials and Methods

## Data Compilation

Since October 2016 through *November 2017*, information was extracted from several data sources both national and international (Figure 1). For each data type, the data was described by 31 categories (e.g. Title, repository, author, spatial and temporal resolution, etc.) creating the metadata (**APENDIX or describing them?**). When possible, all 31 categories were collected. However, in many cases this was not possible. A sub-group of categories were made mandatory for a database to be included on the metadata. These categories were the institution responsible for the data, a short title describing the data, first and last years of data, and a path to find the data or metadata.

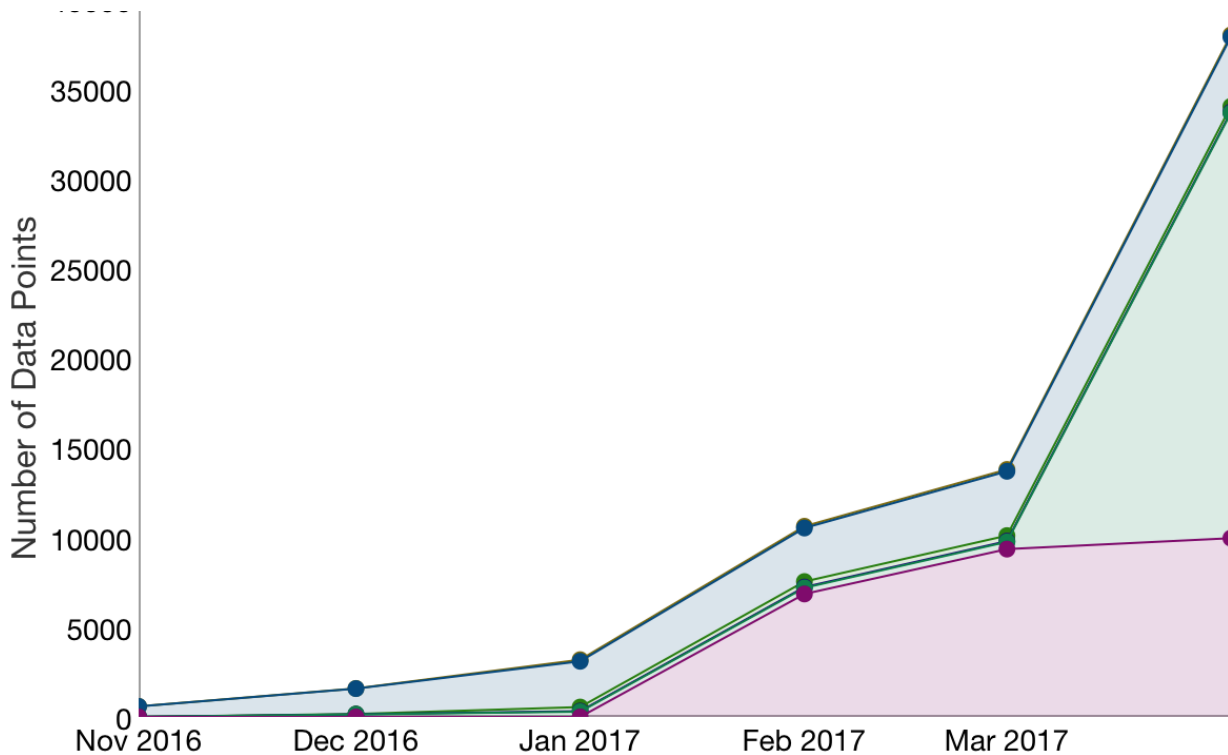


Figure 1: Data Gathering Curve

The data search was divided in three stages (**Table 1**). The first stage target global repositories (e.g. Ocean Bio-geographic Information System (OBIS)), followed by national repositories (e.g. Mexican Fisheries Agency), and regional non-governmental repositories like Monitoreo Noroeste. A second stage target researchers and managers aiming to recover data previously produced but not made available on-line. Professionals were approached at international conferences, workshops, and direct interviews. Finally, a third stage focused the effort on scientific articles and graduate-level thesis. Most Mexican institutions require students to generate a thesis for graduation, sometimes at the undergrad level. Hence, a lot of information that is hardly shared or publicly available is storage in these documents.

Table 1: Number of different sources at each stage of the data gathering process. Note that each source contained multiple data

Stage One	Stage Two	Stage Three
37	40	150

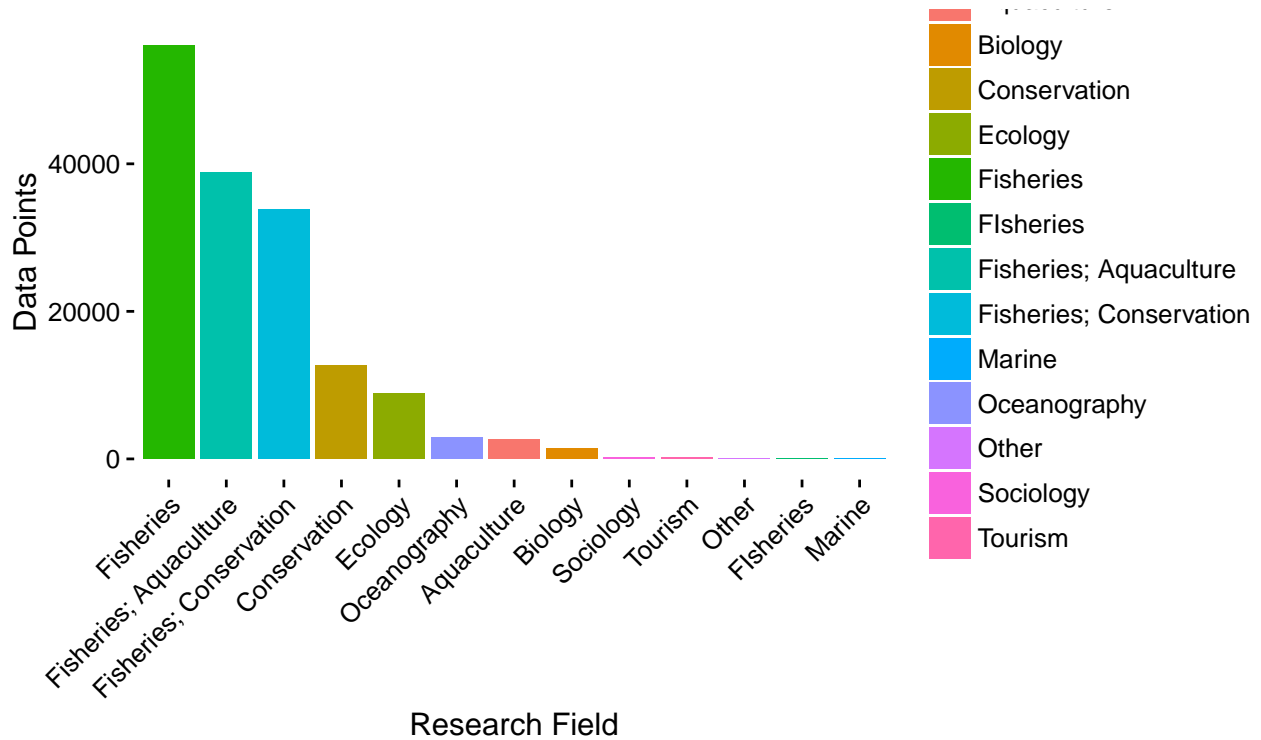


Figure 2: Number of data points per research field.

## Results

### General Numbers

The meta-database has 33053 records containing information for about 157845 data points, from 38 repositories. Data gathered was on 5 major disciplines: Dis from witch *Fisheries* was the most abundant (Figure 2).

### Subject Analysis

There is a total of 7157 subjects in the metadata. The subject that has the most amount of data points is Multiple Species with 32783 units. However, at the species level, *Thunnus albacares* (Albacore tuna) is the most recurrent one with 195 records. The data distribution is very diverse since the mean data point is 18.72(sd 390.42). This is mainly because a large amount of records are one observation of species lists.

### Geographic Analysis

In terms of regional data distribution, data points where evenly distributed between both the Pacific (33818) and the Atlantic oceans (33818). However, the great majority of data points are reported at a National level (27361). At a more detailed level we find that there is a clear dominance of data from both the Gulf of California (15387) and the Campeche Bank (15273), in respect to the other zones determined by the working group.

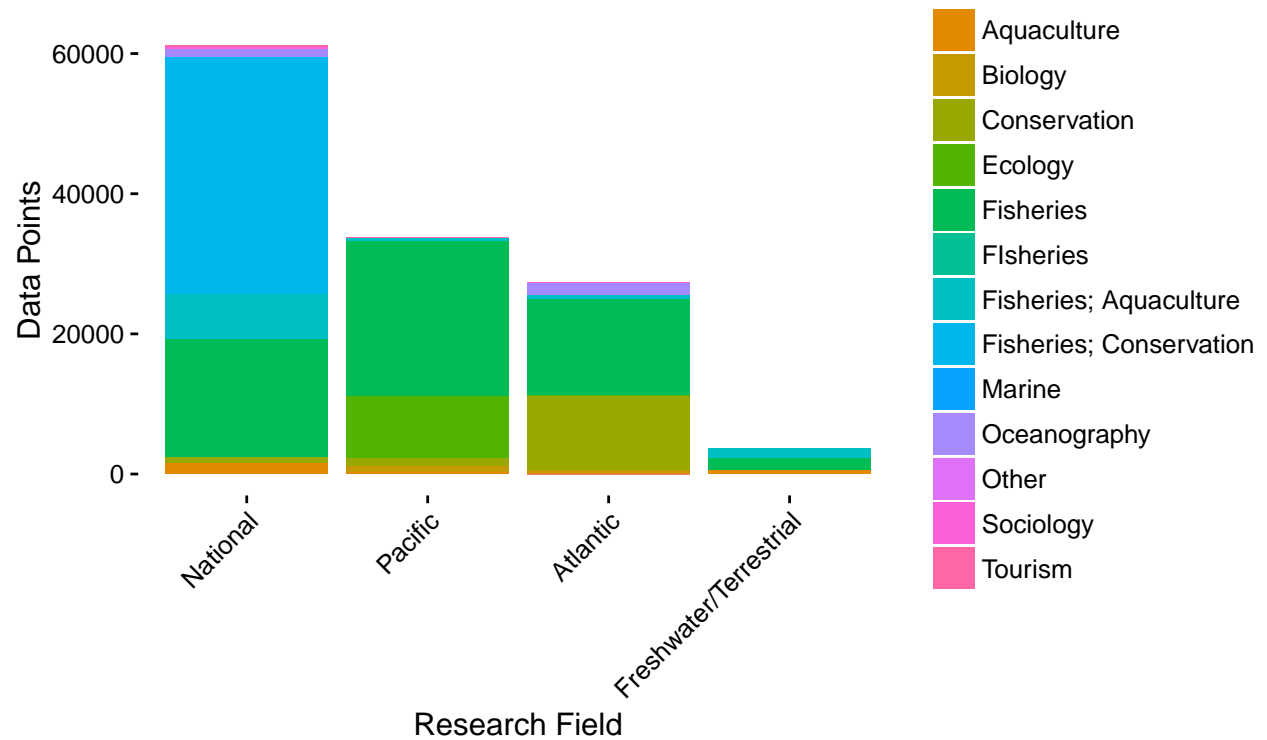


Figure 3: OPTION 1. Number of data points per geographic area for each research field.

## Discussion

## Literature Review