The ocean we want: inclusive and transformative ocean science.
Pelegrí J.L., Gili J.M., Martínez de Albéniz M.V. (eds.). 2022. Institut de Ciències del Mar, CSIC, Barcelona.

6.1. Seventy years of marine science: from notebook to cloud data

Savitri Galiana, Lucía Quirós, Elisa Berdalet, Xavier García, Emilio García-Ladona, Jordi Isern-Fontanet, Laia Viure

Data acquisition, analysis and interpretation are essential tasks for science, business and administration, and research in this field has led to major developments. At the Institut de Ciències del Mar (ICM), researchers have progressed from recording data on paper to storing them on electronic devices or in the "cloud". However, the organization, management and transfer of the data obtained –mostly thanks to public funding– still has many shortcomings. Several initiatives at the ICM are working to overcome these difficulties and organize the scientific data according to the FAIR principles that are applied internationally.

Data storage and management

Before computers, data were recorded on paper and stored in lists, notepads and journals. Texts, charts and tables were drafted by hand and photocopied. Data stored in this way occupied a large physical volume, were slow and difficult to access and were very susceptible to deterioration due to environmental causes and accidental destruction or loss. As computer technology progressed, paper records were first replaced by punched cards that stored the data through a pattern of holes and white spaces, then by magnetic tapes and later by hard disks, floppy disks, CDs, DVDs and flash drives. Recently, thanks to advances in computing and telecommunication technologies, the concept of "cloud data" was introduced. The cloud offers an unlimited amount of data storage capacity and access anytime and anywhere that has an internet connection. The physical warehouse of the cloud consists of a network of servers, often located in different parts of the world, managed by organizations that are responsible for maintaining and protecting the physical system and guaranteeing access to the data.

As technology developed, more advanced modes of data organization and management were introduced, and the concept of databases (DBs) took shape. A database is defined as "a collection of information that is organized so that it can be easily accessed, managed and updated" (Search Data Management Tech Target 2021). Databases are usually controlled through a software known as the database management system (DBMS). A DBMS serves as an interface between the DB and its users or other programs, allowing large amounts of information to be entered, stored, retrieved and organized. The dataset and the DBMS, together with the associated applications, are called the DB system or simply the DB. Databases can be classified according to their organizational model, but the most frequent type are still the relational DBs that appeared in the 1980s. Relational DBs store and organize data in a set of tables with different types of relations between them. However, the most appropriate DB model depends on the type of data to be stored and the reason for storing it. The questions that must be answered in designing a database are i) What is the objective? ii) Who will the users be? and iii) What kind of queries should it answer? In 2016, a consortium of scientists and organizations published the principles that DB must meet for the management of big data (Wilkinson *et al.*

2016): findability, accessibility, interoperability and reusability (FAIR).

The past, present and future of data at the ICM

In the 1970s, the ICM, at that time the Instituto de Investigaciones Pesqueras, carried out the MARESME project, in which the ships R/V *Cornide de Saavedra* and R/V *García del Cid* studied the ocean circulation and characterized in particular the chemical contamination of the coast of Barcelona. Basic hydrographic, chemical and biological observations were made at three fixed oceanographic stations located in a perpendicular section on the coast. Over the course of several days of the year, the scientists sailed there with measuring instruments. Water samples were taken, current meters were installed at fixed points and buoys were thrown to drift with current meters located above and below the thermocline. The data acquired were stored in paper format and transferred through the report, also on paper, *Datos Informativos* 5, "*Datos oceanográficos frente a Barcelona*" (Salat *et al.* 1978). The data were organized in a series of tables, each table corresponding to the measurements made in a certain position and on a certain day (Figure 1). At that time, to analyse the data and be able to find the value of a parameter at a certain depth, position and day, you had to have access to this report, review all the tables,


DATOS INFORMATIVOS

INSTITUTO DE INVESTIGACIONES PESQUERAS

C.S.I.C.

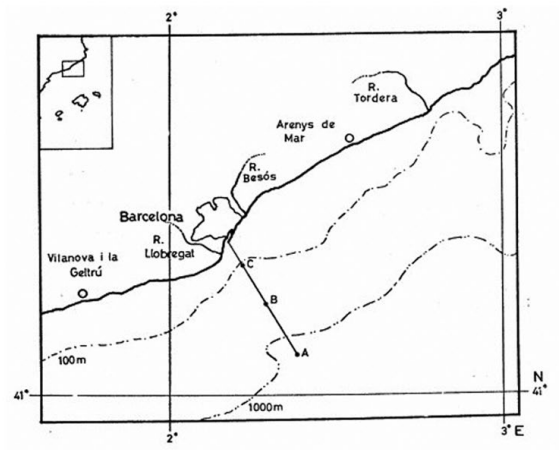
5

Datos oceanográficos frente a Barcelona

(1975 - 1976)

J. SALAT, J. FONT, A. CRUZADO

Barcelona, octubre de 1978



EST NO. 87 LATITUD 41 13.0N D.SECCHI 31 HORAS TIRADAS
FECHA 10/11/75 LONGITUD 2 17.0E PRF.FONDO 350 1 1403 2 1432 3 1451
HORA 1345-1505

PROF	TI	TEMP	SALN	SG-T	VESP	OXIG	NO3	SIO4	PO4	CL-A
0	3	17.94	37.847	27.487	97.325	5.96	.00	.00	.00	
5	2	17.92	37.827	27.477	97.324	6.05	.00	.00	.00	
10	2	17.88	37.843	27.499	97.320	5.93	.00	.00	.00	
25	2	17.20	37.962	27.759	97.289	5.78	.00	.00	.00	
50	2	16.47	38.118	28.055	97.250	5.78	.00	.70	.00	
75	2	15.89	38.104	28.141	97.228	5.80	.45	.70	.09	
100	1	15.09	38.156	28.405	97.197	5.83	1.29	2.60	.09	
150	1	13.37	38.300	28.893	97.129	5.57	4.16	3.70	.27	
200	1	13.24	38.392	28.992	97.099	5.30	5.91	4.70	.17	
250	1	13.26	38.444	29.028	97.074	4.98	6.29	5.90	.23	
300	3	13.67	38.460	28.953	97.061	5.10	7.54	6.40	.32	

Figure 1. An example of storage and transfer of data in the 1970s by the Instituto de Investigaciones Pesqueras, today the Institut de Ciències del Mar. Top left, cover of the report. Top right, location map of the oceanographic stations. Bottom, a data table from station number 87 recorded on 10 November 1975 from the *Datos Informativos* number 5 report of the Instituto de Investigaciones Pesqueras (Salat *et al.* 1978).

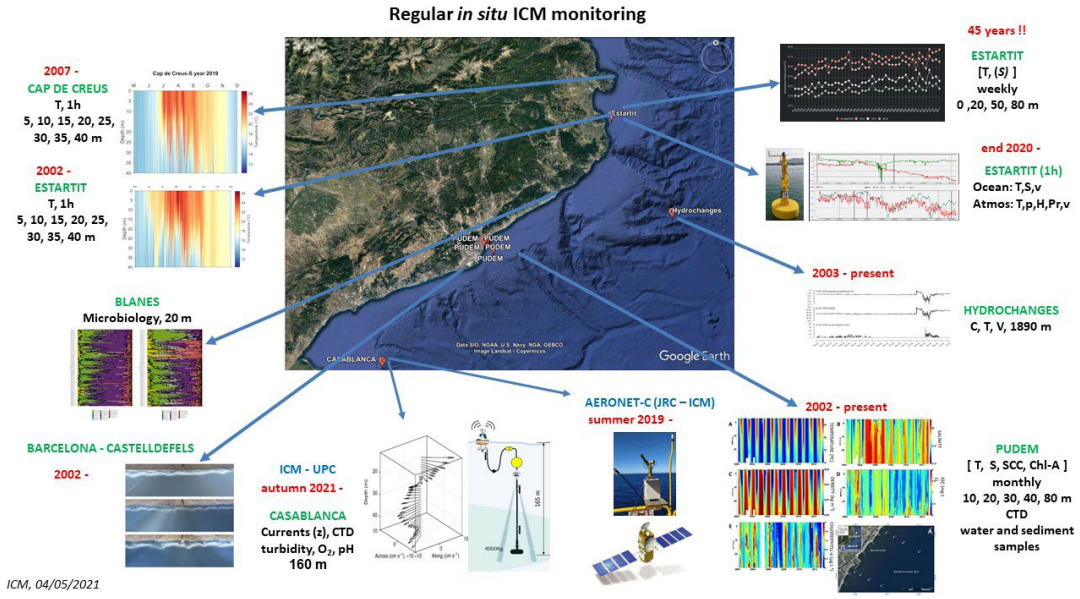


Figure 2. The current panorama of data acquired by the ICM for monitoring the marine environment on the Catalan coast. It is planned to include them in DBs, which will reinforce multidisciplinary studies.

and manually extract the values of interest. Graphs were also drafted manually.

Today, data acquisition during sea-going campaigns, such as Spurs in 2013, has been automated to levels that were unimaginable 40 years ago. The objective of this cruise was to study the oceanographic processes responsible for the formation and maintenance of maximum salinity at the centre of the subtropical shift of the North Atlantic. The researchers used state-of-the-art instrumentation such as drift buoys designed and built at the ICM, which transmitted surface position, temperature and salinity data from the sea via satellite every hour. Some of these buoys were recovered three years later in different parts of the planet, so they were transmitting data in real time every hour for years.

Despite advancements in the data acquisition technology used at the ICM and the possibility of storing data on hard drives, local servers and data clouds, its management and transfer remain quite limited. Most scientists still have their data files stored with specific criteria (typical of the study but not standardized) on their hard drives or servers, without having them organized in a DB. In the coming years, it is expected that more and more machines will deal with the acquisition and processing of data, making marine sciences a discipline based on big data. The European Marine Board Expert Working Group on big data recently launched a series of recommendations to promote the implementation of FAIR principles in the field of marine sciences (Guidi *et al.* 2020). At the ICM, initiatives such as the Xarxa Marítima de Catalunya and IcatMar are developing DBs for fishing and map viewers that allow for better management of fishing resources. In addition, the data capitalization projects SHAREMED and MED OSMoSIS of the Interreg Med programme are working to include data or metadata (descriptive information on the characteristics of the data) of ICM researchers in DBs (Figure 2). This will allow the information to be organized in a much more homogeneous way, guarantee the quality of the data and make them much easier to access, update and transfer. In other words, the FAIR principles will be followed, and this will open up new possibilities for multidisciplinary studies that provide a better understanding of the ocean and its changes.

References

Guidi L., Fernàndez Guerra A., Canchaya C., *et al.* 2020. Big Data in Marine Science. In: Alexander B., Heymans S.J.J., *et al.* (eds), Future Science Brief 6 of the European Marine Board, Ostend, Belgium.

Salat J., Font J., Cruzado A. 1978. Datos oceanográficos frente a Barcelona (1975–1976). Datos Informativos del Instituto de Investigaciones Pesqueras 5:1–73.

Search Data Management Tech Target. 2021. Definition: database. Accessed 17th May 2021. https://searchdatamanagement.techtarget.com/definition/database

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018.

DOI: https://doi.org/10.20350/digitalCSIC/14108