

6.1. De la libreta a la nube de datos: setenta años de ciencia marina

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

La adquisición de datos, análisis y posterior interpretación es una tarea esencial para la ciencia, que requiere almacenar, organizar, acceder y transferir estos datos. Esta necesidad afecta también a otros sectores de la sociedad, como empresas o entidades de gobierno que utilizan los datos como una herramienta clave en la toma de decisiones. Esto ha hecho que la investigación en este campo haya sido muy activa y el almacenamiento y gestión de datos haya sufrido una evolución muy importante. En el Institut de Ciències del Mar (ICM) se ha vivido de cerca este cambio, pasando de tener los datos en papel a tenerlos en dispositivos electrónicos o en la «Nube». Sin embargo, la organización, gestión y transferencia de los datos sigue teniendo muchas carencias. Diferentes iniciativas en el ICM están trabajando para superar las dificultades existentes y organizar la información científica obtenida a lo largo de tantos años, con la financiación de entes públicos fundamentalmente, en bases de datos siguiendo los principios FAIR (*findability, accessibility, interoperability, reusability*) de aplicación internacional en los datos científicos.

Almacenaje y gestión de datos

Antes de que hubiera ordenadores, los datos se registraban en papel y se almacenaban en listas, blocs de notas y revistas, con texto, gráficos y tablas tecleados a mano y ciclostilados. Los datos guardados de esta manera ocupaban mucho volumen físico, su acceso era lento y difícil y eran muy susceptibles a deteriorarse por causas ambientales o a quedar destruidos o extraviados de manera accidental. A medida que

la tecnología fue avanzando, los registros en papel se fueron sustituyendo, primero, por tarjetas perforadas, tarjetas de cartón que almacenaban los datos a través de un patrón de agujeros y espacios blancos; después, por cintas magnéticas y, más adelante, por discos duros, disquetes, CDs, DVDs, memorias USB, etc. Recientemente, gracias a los avances en las tecnologías de la computación y en las telecomunicaciones, se ha extendido el uso de la «Nube» de datos. La Nube ofrece principalmente una cantidad ilimitada de capacidad de almacenamiento de datos y un acceso desde cualquier lugar y en cualquier momento donde haya una conexión *world wide web*. El almacén físico de la Nube consiste en una red de diferentes servidores, muchas veces ubicados en diferentes lugares del mundo, gestionados por organizaciones que se encargan de mantener y proteger el sistema físico, y de garantizar la accesibilidad a los datos.

Con la evolución de las tecnologías se fueron desarrollando modos más avanzados de organización y gestión de los datos, y así fue tomando cuerpo el concepto de las bases de datos (BBDD). Las BBDD se definen como un conjunto de información organizada fácilmente accesible, manipulable y actualizable (Search Data Management Tech Target 2021). Normalmente, se controlan a través de un software conocido como sistema de gestión de BBDD (DBMS, *database management system*). Un DBMS sirve como una interfaz entre la BBDD y sus usuarios u otros programas, permitiendo introducir, almacenar y recuperar grandes cantidades de información, así como de gestionar su organización. El conjunto de datos y el DBMS,

junto con las aplicaciones asociadas, se denominan sistema de BBDD o, simplemente, BBDD. Las BBDD pueden clasificarse en función de su modelo de organización.

A pesar de la fuerte evolución que han experimentado estos modelos, el enfoque más frecuente a día de hoy sigue siendo el de las BBDD relacionales, que aparecieron en la década de los 80. Las BBDD relacionales almacenan y organizan los datos en un conjunto de tablas con diferentes tipos de vinculaciones entre ellas. Sin embargo, el modelo de BBDD más adecuado depende de los propósitos de esta y de los mismos datos. Las preguntas que se deben responder a la hora de diseñar una base de datos son: i) ¿cuál es su objetivo? ii) ¿quiénes serán los usuarios? iii) ¿qué tipo de preguntas debería responder? En 2016, un consorcio de científicos y organizaciones, publicaron los principios que deben cumplir las BBDD (Wilkinson *et al.* 2016): deben poder ser encontradas (*findability*), accesibles (*accessibility*), interoperables (*interoperability*) y reutilizables (*reusability*).

Estos principios se conocen como FAIR por sus siglas en inglés y surgen en el contexto actual de gestión de una cantidad inmensa de datos (*big data*).

Pasado, presente y futuro de los datos en el ICM

En los años 70, el ICM, en aquellos momentos Instituto de Investigaciones Pesqueras, llevó a cabo el proyecto MARESME, en que a bordo del buque B/O Cornide de Saavedra y del B/O García del Cid se estudió la circulación oceánica y se caracterizó la contaminación química de la costa de Barcelona principalmente. Se hicieron observaciones hidrográficas, químicas y biológicas básicas, en tres estaciones oceanográficas fijas situadas en una sección perpendicular a la costa. A lo largo de diferentes días del año, los científicos se trasladaron hasta allí con los instrumentos de medida. Se tomaron muestras de agua, se instalaron correntímetros en puntos fijos y se lanzaron algunas boyas a la deriva con

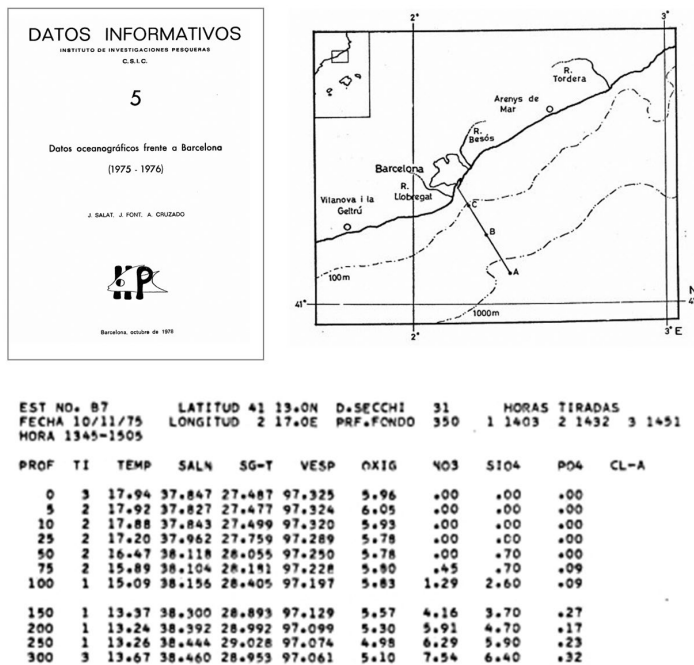


Figura 1. Ejemplo de almacenamiento y transferencia de datos de los años setenta en el Instituto de Investigaciones Pesqueras, hoy, Institut de Ciències del Mar. Portada, mapa de localización de las estaciones oceanográficas y tabla de datos de la estación número 87 registradas el día 10/11/75 del informe Datos Informativos número 5 del Instituto de Investigaciones Pesqueras (Salat *et al.* 1978).

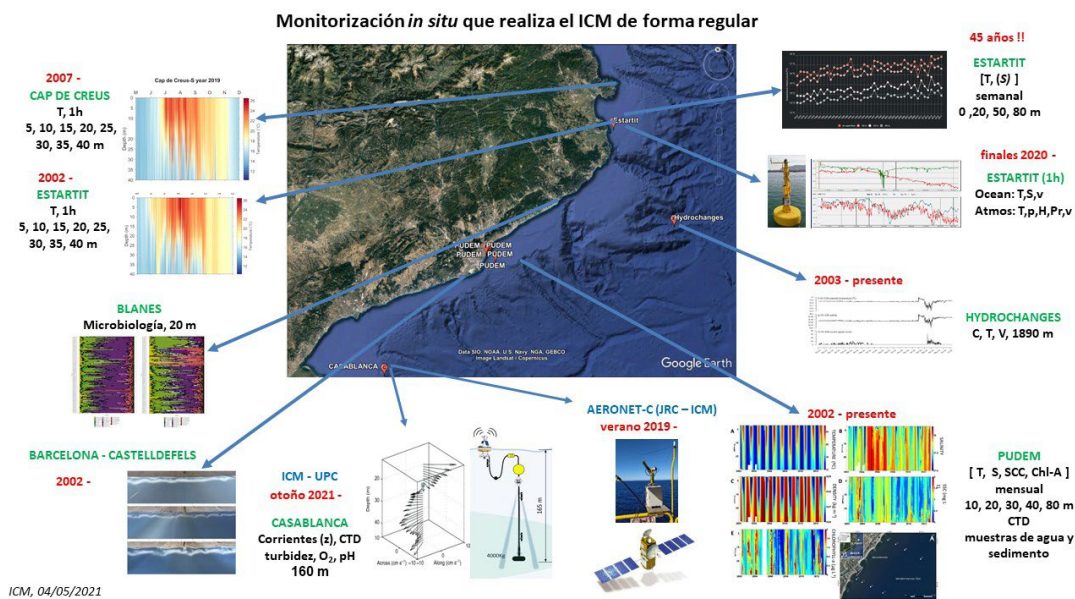


Figura 2. Panorama actual de diferentes datos que adquiere el ICM para la monitorización del medio marino en la costa catalana. Está previsto de incluirla en BBDD, lo que reforzará los estudios multidisciplinares..

correntímetros situados por encima y por debajo de la termoclina. Los datos adquiridos se almacenaron en formato papel y se transfirieron a través del informe, también en papel, Datos Informativos 5, «Datos oceanográficos frente a Barcelona» (Salat *et al.* 1978), organizados en una serie de tablas, donde cada tabla correspondía a las medidas hechas en una posición y un día determinados (figura 1). En aquellos tiempos, para analizar los datos y poder encontrar el valor de un parámetro a una cierta profundidad, posición y día, había que tener acceso a este informe, revisar todas las tablas y manualmente extraer el valor de interés. Las gráficas se hacían también manualmente.

Hoy día, la adquisición de datos en campañas, como, por ejemplo, la SPURS de 2013, se ha automatizado hasta niveles inimaginables cuarenta años atrás. El objetivo de esta campaña era el estudio de los procesos oceanográficos responsables de la formación y mantenimiento del máximo de salinidad en el centro del giro subtropical del Atlántico Norte. Para ello, se utilizó instrumentación de última generación, como, por ejemplo, unas boyas de deriva diseñadas y construidas en el ICM, que transmitían cada hora datos de posición, temperatura y salinidad superficiales del

mar vía satélite. Algunas de estas boyas se recuperaron tres años más tarde en diferentes lugares del planeta, por lo que estuvieron transmitiendo datos en tiempo real cada hora, durante años.

A pesar del avance en la tecnología de adquisición de datos que se utiliza en el ICM, y la posibilidad de guardar los datos en discos duros, servidores locales y Nubes de datos, su gestión y transferencia sigue siendo bastante limitada. La mayoría de los científicos siguen teniendo sus ficheros de datos almacenados con criterios específicos (propios del estudio, pero no estandarizados), en sus discos duros, servidores, etc. sin tenerlos organizados en BBDD. Por otra parte, en los próximos años, se espera que sean cada vez más las máquinas que se ocupen de la adquisición y procesado de datos, haciendo de las ciencias marinas una disciplina basada en el *big data*. En este contexto, recientemente, el European Marine Board Expert Working Group en *big data* ha lanzado una serie de recomendaciones para fomentar la implementación de los principios FAIR en el campo de las ciencias marinas (Guidi *et al.* 2020). En el ICM, iniciativas como la *Xarxa Marítima de Catalunya* y el IcatMar, están desarrollando BBDD de pesca y visualizadores espaciales que permitan una mejor gestión de los recursos pes-

queros. Además, los proyectos de capitalización de datos, SHAREMED y MED OSMoSIS, del Interreg Mediterranean program, están trabajando para incluir los datos o metadatos (información descriptiva sobre las características de los datos) de aquellos investigadores del ICM que lo deseen en BBDD (figura 2). Esto permitirá organizar la información de una manera mucho más homogénea y garantizar la calidad de los datos, será mucho más fácil acceder a ellos, se podrán actualizar y serán fácilmente transferibles. Es decir, seguirán los principios FAIR y, además, abrirán nuevas posibilidades a estudios multidisciplinares que permitan una mejor comprensión del océano y sus cambios.

Referencias

- Guidi L., Fernández 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>