

6.1. De la llibreta al núvol de dades: 70 anys de ciència marina

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

L'adquisició de dades, anàlisi i posterior interpretació és una tasca essencial per a la ciència, que requereix emmagatzemar, organitzar, accedir i transferir aquestes dades. Aquesta necessitat afecta també altres sectors de la societat, com empreses o entitats de govern que utilitzen les dades com una eina clau en la presa de decisions. Això ha fet que la recerca en aquest camp hagi estat molt activa i l'emmagatzematge i gestió de dades hagi sofert una evolució molt important. A l'Institut de Ciències del Mar (ICM) s'ha viscut de prop aquest canvi, passant de tenir les dades en paper a tenir-les en dispositius electrònics o al «Núvol». Tot i així, l'organització, gestió i transferència de les dades segueix tenint moltes mancances. Diferents iniciatives a l'ICM estan treballant per a superar les dificultats existents i organitzar la informació científica obtinguda al llarg de tants anys, amb el finançament d'ens públics fonamentalment, en bases de dades seguint els principis FAIR (findability, accessibility, interoperability, reusability) d'aplicació internacional a les dades científiques.

Emmagatzematge i gestió de dades

Abans no hi haguessin ordinadors, les dades s'enregistraven en paper i s'emmagatzemaven en llistes, blocs de notes i revistes, amb text, gràfics i taules teclejats a mà i ciclostilats. Les dades guardades d'aquesta manera ocupaven molt volum físic, el seu accés era lent i dificultós i eren molt susceptibles a deteriorar-se per causes ambientals o a quedar destruïdes o extraviades de manera accidental. A mesura que la tecnologia va anar avançant, els registres en paper es van

anar substituint, primer, per targetes perforades, targetes de cartró que emmagatzemaven les dades a través d'un patró de forats i espais blancs; després, per cintes magnètiques i, més endavant, per discs durs, disquets, CDs, DVDs, memòries USB, etc. Recentment, gràcies als avenços en les tecnologies de la computació i en les telecomunicacions, s'ha estès l'ús del «Núvol» de dades. El Núvol ofereix principalment una quantitat il·limitada de capacitat d'emmagatzematge de dades i un accés des de qualsevol lloc i en qualsevol moment on hi hagi una connexió *world wide web*. El magatzem físic del Núvol consisteix en una xarxa de diferents servidors, molts cops ubicats en diferents indrets del món, gestionats per organitzacions que s'encarreguen de mantenir i protegir el sistema físic, i de garantir l'accessibilitat a les dades.

Amb l'evolució de les tecnologies es van anar desenvolupant modes més avançats d'organització i gestió de les dades, i així va anar prenent cos el concepte de les bases de dades (BBDD). Les BBDD es defineixen com un conjunt d'informació organitzada fàcilment accessible, manipulable i actualitzable (Search Data Management Tech Target 2021). Normalment, es controlen a través d'un software conegut com a sistema de gestió de BBDD (DBMS, database management system). Un DBMS serveix com una interfície entre la BBDD i els seus usuaris o altres programes, permetent d'introduir, emmagatzemar i recuperar grans quantitats d'informació, així com de gestionar-ne l'organització. El conjunt de dades i el DBMS, juntament amb les aplicacions associades, s'anomenen sistema de BBDD o, simplement, BBDD. Les BBDD

poden classificar-se en funció del seu model d'organització. Tot i la forta evolució que han experimentat aquests models, l'enfocament més freqüent a dia d'avui segueix sent el de les BBDD relacionals, que aparegueren a la dècada dels 80. Les BBDD relacionals emmagatzemen i organitzen les dades en un conjunt de taules amb diferents tipus de vinculacions entre elles. Tot i així, el model de BBDD més adient depèn dels propòsits d'aquesta i de les mateixes dades. Les preguntes que s'han de respondre a l'hora de dissenyar una base de dades són: i) quin n'és l'objectiu? ii) qui en seran els usuaris? iii) quin tipus de preguntes hauria de respondre? El 2016, un consorci de científics i organitzacions van publicar els principis que han de complir les BBDD (Wilkinson *et al.* 2016): han de poder ser trobades (*findability*), accessibles (*accessibility*), interoperables (*interoperability*) i reutilitzables (*reusability*). Aquests principis es coneixen com a FAIR per les sigles en anglès i sorgeixen en el context actual de gestió d'una quantitat immensa de dades (*big data*).

Passat, present i futur de les dades a l'ICM

Als anys 70, l'ICM, en aquells moments Institut de Investigaciones Pesqueras, va dur a terme el projecte MARESME, en què a bord del V/O Cornide de Saavedra i del V/O García del Cid es va estudiar la circulació oceànica i es va caracteritzar la contaminació química de la costa de Barcelona principalment. S'hi van fer observacions hidrogràfiques, químiques i biològiques bàsiques, en tres estacions oceanogràfiques fixes situades en una secció perpendicular a la costa. Al llarg de diferents dies de l'any, els científics es van traslladar fins allà amb els instruments de mesura. Es van agafar mostres d'aigua, es van instal·lar correntòmetres en punts fixos i es van llençar algunes boies a la deriva amb correntòmetres situats per sobre i per sota de la termoclina. Les dades adquirides es van emmagatzemar en format paper i es van transferir a través de l'informe, també en paper, *Datos Informativos 5*, «*Datos oceanográficos frente a Barcelona*» (Salat

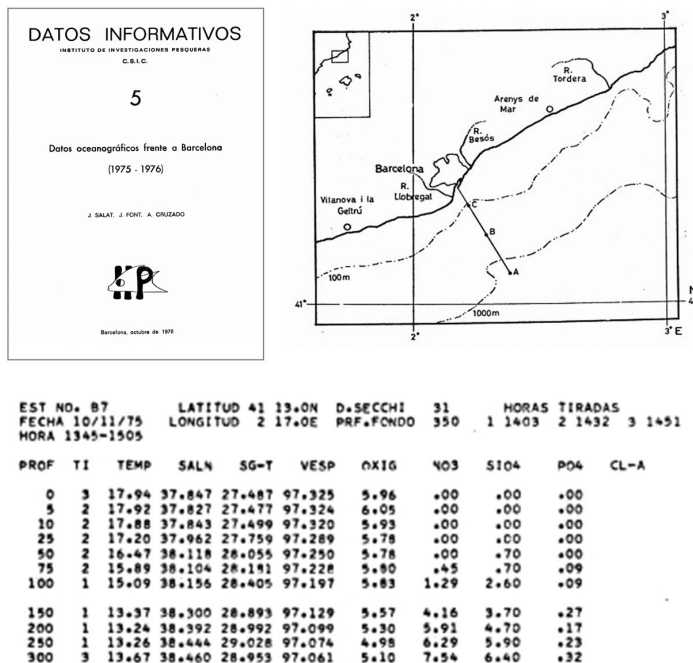


Figura 1. Exemple d'emmagatzematge i transferència de dades dels anys 70 a l'Institut de Investigaciones Pesqueras, avui, Institut de Ciències del Mar. Portada, mapa de localització de les estacions oceanogràfiques i taula de dades de l'estació número 87 enregistrades el dia 10/11/75 de l'informe *Datos Informativos* número 5 de l'Institut de Investigaciones Pesqueras (Salat *et al.* 1978).

Mesures de seguiment del medi marí enregistrades *in situ* de forma regular per l'ICM

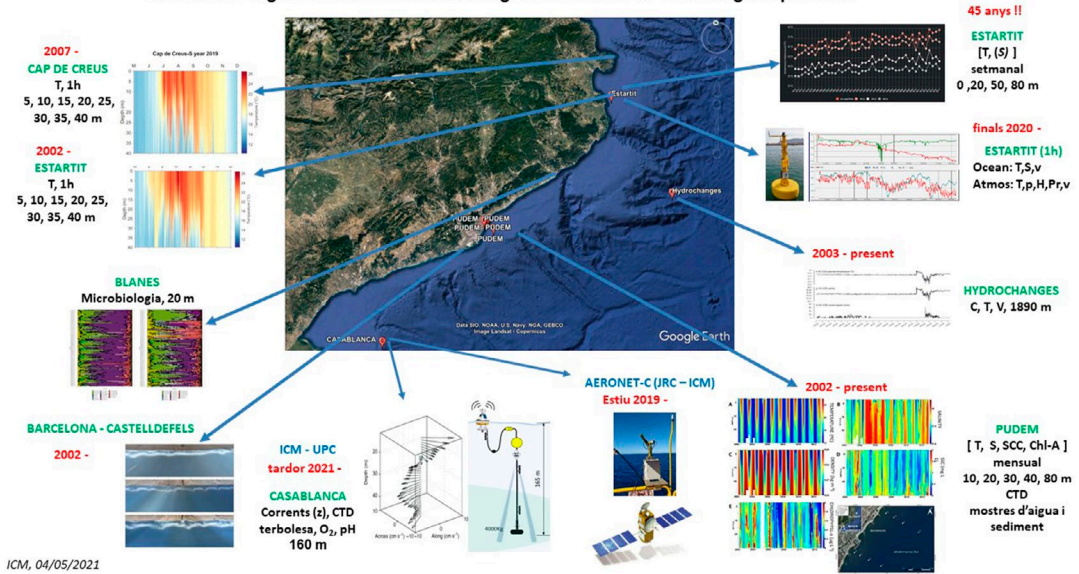


Figura 2. Panorama actual de diferents dades que adquireix l'ICM per al seguiment del medi marí a la costa catalana. Està previst d'incloure-les en BBDD, la qual cosa reforçarà els estudis multidisciplinaris.

et al. 1978), organitzades en una sèrie de taules, on cada taula corresponia a les mesures fetes en una posició i un dia determinats (figura 1). En aquells temps, per analitzar les dades i poder trobar el valor d'un paràmetre a una certa profunditat, posició i dia, s'havia de tenir accés a aquest informe, revisar totes les taules i manualment extreure el valor d'interès. Les gràfiques es feien també manualment.

Avui dia, l'adquisició de dades en campanyes, com per exemple, la SPURS del 2013, s'ha automatitzat fins a nivells inimaginables 40 anys enrere. L'objectiu d'aquesta campanya era l'estudi dels processos oceanogràfics responsables de la formació i manteniment del màxim de salinitat en el centre del gir subtropical de l'Atlàntic Nord. Per a això, es va utilitzar instrumentació d'última generació, com per exemple, unes boies de deriva dissenyades i construïdes a l'ICM, que transmetien cada hora dades de posició, temperatura i salinitat superficials del mar via satèl·lit. Algunes d'aquestes boies es van recuperar tres anys més tard a diferents indrets del planeta, pel que van estar transmetent dades en temps real cada hora, durant anys.

Tot i l'avenç en la tecnologia d'adquisició de dades que s'utilitza a l'ICM, i la possibilitat de guardar les dades en discs durs, servidors locals i Núvols de dades, la seva gestió i transferència segueix essent força limitada. La majoria dels científics segueixen tenint els seus fitxers de dades emmagatzemats amb criteris específics (propis de l'estudi, però no estandaritzats), en el seus discs durs, servidors, etc. sense tenir-los organitzats en BBDD. D'altra banda, en els propers anys, s'espera que siguin cada cop més les màquines qui s'ocupin de l'adquisició i processament de dades, fent de les ciències marines una disciplina basada en el *big data*. En aquest context, recentment, l'European Marine Board Expert Working Group en *big data* ha llençat una sèrie de recomanacions per fomentar la implementació dels principis FAIR en el camp de les ciències marines (Guidi *et al.* 2020). A l'ICM, iniciatives com la Xarxa Marítima de Catalunya i l'IcatMar, estan desenvolupant BBDD de pesca i visualitzadors espacials que permeten una millor gestió dels recursos pesquers. A més a més, els projectes de capitalització de dades, SHAREMED i MED OSMoSIS, de l'Interreg Mediterranean program, estan treballant per

incloure les dades o metadades (informació descriptiva sobre les característiques de les dades) d'aquells investigadors de l'ICM que ho vulguin en BBDD (figura 2). Això permetrà organitzar la informació d'una manera molt més homogènia i garantir la qualitat de les dades, serà molt més fàcil accedir-hi, es podran actualitzar i seran fàcilment transferibles. És a dir, seguiran els principis FAIR i a més a més, obriran noves possibilitats a estudis multidisciplinaris que permetin una millor comprensió de l'oceà i els seus canvis.

Referències

- 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., *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>