

# IMPORTANCIA DE LA BIG DATA EN LAS CIENCIAS MARINAS

VILLARROEL JESÚS<sup>1\*</sup> & MIZRACHI GETRUDIS<sup>2</sup>

*Universidad de oriente, Núcleo Nueva Esparta. Escuela de Ciencias Aplicadas del Mar (ECAM). Departamento de Acuicultura, Boca de Río, Isla de Margarita, Venezuela.*  
*<sup>1\*</sup> autor de correspondencia: Jesusvil281197@gmail.com; <sup>2</sup>gmizrachi2007@gmail.com*

**RESUMEN:** El término Big Data hace referencia al conjunto de datos que por sus características no pueden ser procesados usando herramientas convencionales. Sus principales características son: volumen, velocidad, variedad, valor y veracidad. El término fue empleado por primera vez en el año 1997, y desde entonces ha tenido gran importancia en varios campos de la sociedad, incluyendo las ciencias marinas. La cadena de valor de Big Data es una técnica que permite analizar y aprovechar la información obtenida de la Big Data en la toma de decisiones acertadas, y se compone de tres pasos: adquisición, análisis y curación de datos. Las fuentes de datos en las ciencias marinas provienen de instituciones gubernamentales como la NOAA, NASA y NODC y varios dispositivos empleados para la colecta de datos. La oceanografía, la biología marina, la biología pesquera y la acuicultura son las principales ciencias que se verán beneficiadas por el auge del Big Data. En la actualidad existen importantes proyectos de Big Data en las ciencias marinas, entre los cuales se pueden destacar "Real Time Aquaculture" desarrollado por IBM Deep Sense y Fishtalk Control, por Akva Group, y sus actualizaciones, orientados a facilitar el monitoreo de las granjas de cultivo de peces y obtener información en tiempo real de las condiciones de los cultivos, así como información de finanzas y mercadeo.

*Palabras clave: Big Data, Ciencias marinas, importancia,*

**ABSTRACT:** The term Big Data refers to the set of data that, due to its characteristics, cannot be processed using conventional tools. Its main characteristics are volume, velocity, variety, value and veracity. The term was first used in 1997, and since then it has been of great importance in various fields of society, including marine sciences. The Big Data value chain is a technique for analyzing and leveraging information obtained from Big Data in making sound decisions, and is composed of three steps: data acquisition, analysis, and curation. Data sources in marine sciences come from government institutions such as NOAA, NASA and NODC and various devices used for data collection. Oceanography, marine biology, fisheries biology and aquaculture are the main sciences that will benefit from the rise of Big Data. Currently there are important Big Data projects in marine sciences, among which we can highlight "Real Time Aquaculture" developed by IBM Deep Sense and Fishtalk Control, by Akva Group, and its updates, aimed at facilitating the monitoring of fish farming farms and obtaining real-time information on the conditions of the crops as well as financial and marketing information.

*Keywords: Big Data, Marine Sciences, Importance*

## 1.- INTRODUCCIÓN

Los seres humanos estamos generando y almacenando información (MORENO-CARRILES 2018). A este conjunto enorme de datos contenidos en la red se le denomina Big Data y es un término que hace referencia a un conjunto de datos que, por su volumen, variedad de fuentes y velocidad de producción no pueden ser analizados utilizando procesos o herramientas

tradicionales. El término Big Data fue utilizado por primera vez en julio de 1997 por MICHEL COX Y DAVID ELLESWORTH (COX & ELLESWORTH, 1977) de la NASA. Ellos previeron las dificultades que tendrían los sistemas informáticos de la época para manejar un incremento en el ritmo de crecimiento del volumen de datos. Desde entonces el Big Data ha tomado parte en los distintos sectores de la sociedad humana. A lo largo de la última década el concepto de Big Data se ha ido transformando hasta llegar a un nuevo concepto, el “Data Science”, que es una forma de referirse al conjunto de técnicas empleadas para tratar y manejar datos masivos. E incluso de este ha surgido un nuevo perfil profesional, llamado “Data Scientist”, profesionales que conocen las herramientas computacionales y estadísticas usadas para el tratamiento y manejo de datos (HERNÁNDEZ-LEAL *et al.* 2017). Actualmente, Big Data incluye una amplia diversidad de tipos de datos (no solo números, sino también imágenes, audios y videos), sin embargo, todos ellos tienen 5 características principales, conocidas como las 5 Vs: Velocidad, Volumen, Variedad, Veracidad y Valor. Ahora bien, existen muchos dominios donde las aplicaciones de la Big Data continúan sin desarrollarse. El océano, los mares y áreas marinas cubren cerca de un 72% del planeta y cerca del 80% es aún inexplorada (NATIONAL GEOGRAPHIC 2022). Se prevé que la llamada economía azul, aunque difícil de monetizar, se convierta en una industria muy importante que se espera tenga un crecimiento incluso mayor en los próximos años (LYTRA *et al.* 2017). En este orden de ideas, el Big Data Marino lo constituyen todos los datos obtenidos en el ámbito marino. Estos pueden obtenerse mediante el uso de los sensores de distintos dispositivos, por ejemplo, aquellos que sean capaces de registrar información sobre posicionamiento global (como GPS) junto con mediciones de las condiciones ambientales del agua (LYTRA *et al.* 2017). En consecuencia y siendo, en líneas generales, tan alto el porcentaje de datos desconocidos sobre el océano, aunado a la gran cantidad de actividades que se desarrollan en torno él; que se piensa que, en las próximas décadas, este Big Data Marino tendrá un crecimiento exponencial. Por tal motivo, se hace necesario entender el posible impacto de esta producción de datos en las ciencias Marinas. En base a lo anterior se estableció el objetivo de investigar la importancia de la Big Data en las Ciencias Marinas.

## 2.- DESARROLLO

### 2.1.- Fuentes de Datos en las Ciencias Marinas

La Big Data no se considera una tecnología concreta sino una situación en la que las necesidades de almacenamiento y procesamiento de información tanto estructurada como no estructurada sobrepasan las capacidades de las tecnologías tradicionales. Estos datos pueden obtenerse mediante sensores. Estos son instalados en la embarcación, y durante todos sus viajes y trayectos recolectan y almacenan grandes volúmenes de datos *in situ* describiendo las condiciones de la embarcación en tiempo real, así como el estado de las comunicaciones terrestres, las condiciones climáticas y las condiciones oceánicas (LYTRA *et al.* 2017). No solo sensores en buques, sino también satélites (FABLET *et al.* 2017), boyas (IVANOVA & SAHOY 2017) y vehículos autónomos (HOLLAND *et al.*, 2016). Estos datos son proporcionados por organizaciones gubernamentales y científicas como la NOAA (National Oceanic and Atmospheric Administration), NASA (National Aeronautics and Space Administration) y NODC (National Oceanographic Data Collection), mediante infraestructuras web integradas en las que están disponibles los datos, metadatos y otros productos en el ambiente marino.

## **2.2.- Las 5 vs del Big Data Marino**

Según LYTRA *et al.* (2017) el Big Data marino se caracteriza por:

Volumen: existen fuentes de datos marinos que proporcionan una enorme cantidad de estos. Por ejemplo, están las páginas web de organizaciones científicas como NOAA, NASA y NODC que coleccionan terabytes de información por día y permiten acceder a un vasto número de conjuntos de datos de biodiversidad marinos y oceanográficos. Velocidad: consiste en la velocidad de recolección de datos. Organizaciones como la NOAA ofrecen condiciones colectadas en vivo cada media hora vía satélites del GOES y comerciales. Se estiman que mensualmente recolectan más de 26 terabytes de datos oceánicos, atmosféricos y geofísicos. Variación: se refiere a la variabilidad que existe entre los métodos de recolección, instrumentos, grados de precisión, tipos, fuentes de datos y su manera de ser almacenados en diferentes formatos y estructuras. Valor: La Big Data Marina puede ser utilizada para mejorar el valor del manejo de datos de las aplicaciones marinas. Por ejemplo, los cruceros oceanográficos y la data meteorológica representan la base para la predicción de operaciones técnicas y mantenimiento de timones y estaciones. Veracidad: Debido a que la data es colectada siguiendo una amplia variedad de métodos y usando disparidad de instrumentos, diferentes grados de precisión y errores son observados en los datos de fuentes marinas.

## **2.3.- Cadena de valor del Big Data**

Esta cadena es el centro del futuro conocimiento económico brindando oportunidades de desarrollo digitales a sectores tradicionales de la Big Data públicas y privadas, en asociación. Esta se compone de una serie de pasos que son los siguientes: El primer paso de la cadena de valor de Big Data es la adquisición. Durante esta fase, se emplean métodos para extraer datos crudos desde diferentes fuentes y formatos y se organizan. Una vez hecho esto, se procede a limpiar y preparar los datos para depurarlos. La siguiente etapa es el análisis de los datos por medio de análisis exploratorios, gráficos y modelación, un proceso que busca extraer información de los datos almacenados para poder generar y facilitar la toma de decisiones (HERNÁNDEZ-LEAL *et al.* 2017). Durante la depuración o curación de datos, se juzga la calidad de los mismos y se integran. Para ello se emplean diferentes métodos de compresión de datos, indexación y acceso para almacenar efectivamente los datos. Finalmente, su valor y veracidad dependerá de los patrones y predicciones que se hagan durante la fase de procesado y análisis.

## **2.4.- Ciencias marinas impulsadas por Big Data**

Una de las ciencias marinas que recibirá el impulso directo del Big Data es la oceanografía y sus especialidades, ya que nos permite comprender los impactos del cambio climático en el océano (GUIDI *et al.* 2020). Recientes adiciones al paquete de datos de Python, como Xarray y Dask, proporcionan a los científicos la posibilidad de aplicar computación paralela para trabajar con datos que por su gran volumen no pueden almacenarse en la memoria física local de los computadores ordinarios, tales como los sets de datos asociados al área de Big Data. La potencia con la que los sistemas producen datos es igual a la cantidad de datos oceánicos por observaciones multiplicadas por un factor de 100. Esta cantidad se incrementa con el tiempo, así como la resolución del “ocean ice” y la capacidad de las computadoras, No obstante, la poca cantidad de

datos observables ha sido un desafío para la oceanografía, debido a la opacidad del océano ante mediciones por radiación electromagnética y altos precios de subirse a una embarcación para tomar muestras in situ. Recientemente ha habido una explosión tanto en la riqueza de los datos como en nuevas iniciativas como un teleférico submarino, y plataformas autónomas con sensores para solventar este problema. Actualmente programas como ARGO, The International Global Drifter Program y la Ocean Observatories Initiative están produciendo una cantidad de datos in situ sin precedentes y con alta resolución. Esta iniciativa propuso la instalación de una red de instrumentos oceanográficos en siete lugares alrededor del mundo. Estos equipos se componen de boyas, cámaras submarinas y vehículos autónomos equipados con miles de sensores que colectan datos acerca de la temperatura del agua, salinidad, niveles de clorofila, actividad volcánica y mucho más en diferentes regiones del océano, como en la superficie, la columna de agua y el suelo oceánico. Estos dispositivos, a su vez, son portátiles y en un futuro pueden ser trasladados de un lugar a otro. Y gracias a sus sensores los investigadores pueden monitorear el afloramiento del plancton y establecer sus causas mediante modelaciones estadísticas, comprender las pesquerías y los stocks pesqueros y cómo reacciona el océano a la actividad humana. Pero no solo eso, también nos permitiría estudiar el océano profundo y su influencia en la superficie, como en el caso de las fosas hidrotermales, la principal fuente de minerales del océano y facilitar el estudio del océano como un sistema íntegro y no como conjunto de ambientes fragmentados. Otros problemas planteados en la oceanografía son la introducción de datos de gran escala y la reproductibilidad de los datos en computadoras. Uno de los proyectos de Big Data en oceanografía más importantes es el Big Data Program (BDP) desarrollado por la NOAA en 2015 (NOAA) en conjunto con otras empresas en un acuerdo no competitivo llamado CRADA's. Este programa es una combinación de tres recursos increíblemente poderosos: la gran y costosa cantidad de datos ambientales y experiencia de la NOAA, la capacidad de cálculo escalable de las tecnologías de sus colaboradores y la energía e innovación de la economía. Este programa almacena los datos en plataformas modernas con servicios de alto rendimiento, almacenaje y análisis que permitieron crear un espacio económico y expandir estos datos hacia el público en general. Uno de esos servicios es el Cloud Service Providers (CSP's), un servicio a donde toda la data se carga mediante un Data broker para su uso por parte del público general de forma gratuita y sin restricciones de uso. No obstante, cualquier uso de datos con valor agregado puede incurrir en un cargo. Por otra parte, y además de la oceanografía, otras de las ciencias que se verán beneficiadas por el auge del Big Data son la biología marina, pesquera y la acuicultura. A diferencia de la oceanografía, estas ciencias emplean las propiedades biológicas para mejorar nuestro entendimiento de los ecosistemas marinos y la biodiversidad bajo el concepto del cambio constante. La medición de los parámetros de estas ciencias representa un desafío y su logro habilitaría productos de data científica relacionados a diferentes legislaciones para el manejo de la biodiversidad marina. Un proyecto perteneciente a la IBM DeepSense con "Real Time Aquaculture" permite el monitoreo impulsado por datos de granjas de cultivo en los océanos utilizando sensores inalámbricos que envían datos a cada minuto, facilitando la toma de decisiones de los grajeros (AYMAN 2008). Estos sensores toman 100,000 mediciones diarias y analizan 11 millones de datos relacionados a parámetros fisicoquímicos de la granja. Otro proyecto a destacar es Fishtalk Control, una herramienta desarrollada por Akva Group Chile que permite capturar información de las granjas salmoneras desde la reproducción hasta la cosecha a través de

distintos módulos, incluyendo información relacionada con finanzas y planificación, lo cual permite tomar decisiones en base a biología y economía (AKVA GROUP). De entre estos se puede mencionar el llamado Fishtalk Optimazer, que analiza las condiciones generales del mercado para obtener el máximo rendimiento de biomasa. Esta herramienta está modernizándose permanentemente y recientemente han sacado nuevas versiones de la misma, como Fishtalk Plan, enfocada hacia los requerimientos de los clientes y Fishtalk App, dedicada a facilitar la introducción de datos en la base de datos a través de distintos dispositivos móviles e incluso off line.

### 3.- CONCLUSIONES

**3.1.-** La Big Data Marina es uno de los campos con mayor producción y complejidad de datos de la próxima década. **3.2.-** Dentro de la Big Data Marina, las Ciencias Marinas recibirán un impacto significativo en cuanto a adquisición, volumen, manejo, procesamiento, servicio, interoperabilidad y accesibilidad de los datos. **3.3.-** Con la finalidad de realizar un manejo eficiente se impulsará el desarrollo de la arquitectura de datos y su interacción con los diferentes proyectos que evolucionen dentro del área marina.

### 4.- REFERENCIAS

- AKVA GROUP. *FISHTALK CONTROL: desde la producción hasta la cosecha*. Disponible en <https://www.akvagroup.com/software-s/fishtalk-control>
- AYMAN, A. 2008. *IBM DeepSense Collaboration to advance in Blue Economy in Canada*. IBM Think Blog. Disponible en: <https://www.ibm.com/blogs/think/2018/09/ibm-deepsense/>
- COX, M. & ELLSWORTH, D. 1997. Managing big data for scientific visualization. *ACM siggraph*. 97 (1): 21-38.
- FABLET, R., P. H. VIET, R. LGUENSAT & B. CHAPRON. 2017. Exploiting ocean observation and simulation big data to improve satellite-derived geophysical products: Analog strategies. *Proc. Big Data Space Conf*: 169-172.
- GUIDI, L., A. FERNÁNDEZ GUERRA, C. CANCHAYA, E. CURRY, F. FOGLINI, J-O IRISSON, K. MALDE, C. T. MARSHALL, M. OBST, R. P. RIBEIRO, J. TJIPUTRA & D. C. BAKKER. 2020. *Big Data in Marine Science*. Future Science Brief of the European Marine Board, Ostend, Belgium. 50 pp.
- HOLLAND, M., A. HOGGARTH & J. NICHOLSON. 2016. Hydrographic processing considerations in the “Big Data” age: An overview of technology trends in ocean and coastal surveys. *IOP Conf. Ser.: Earth Environ. Sci.* 34(1).
- HERNÁNDEZ-LEAL, E.J., N. D. DUQUE-MÉNDEZ & J. MORENO-CADAVID. 2017. *Big Data: una exploración de investigaciones, tecnologías y casos de aplicación*. *TecnoL*. 20 (39):17-24.
- IVANOVA, D. & S. ZAHOV. 2017. Ocean Big Data Processing in IoT Ecosystem.
- LYTRA, I., M. E. VIDAL, F. ORLANDI & J. ATTARD. 2017. A Big Data Architecture for Managing Oceans of Data and Marine Applications. *International Conference on Engineering, Technology and Innovation*: 1216-1226.

MORENO-CARRILES, R.M. 2018. *Big data, ¿pero ¿qué es?* *ANGIO*. 650(1): 1-4.

NATIONAL GEOGRAPHIC. 2022. Resource Library: Ocean. Disponible en: <https://education.nationalgeographic.org/resource/ocean/#:~:text=Despite%20its%20size%20and%20impact,of%20our%20own%20ocean%20floor>.

National Oceanic and Atmospheric Organization (NOAA). *Big Data Program*. Disponible en: <https://www.noaa.gov/organization/information-technology/big-data-program>.

RECIBIDO: AGOSTO 2022

ACEPTADO: NOVIEMBRE 2022