



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE MÁSTER

*Gemelo Digital de Entorno Operativo Marítimo: Propuesta de
arquitectura de integración de datos y operación*

Máster Universitario en Dirección TIC para la Defensa

ALUMNO: Sergio Ramírez Morán

DIRECTORES: Milagros Fernández Gavilanes
Carlos Pérez Collazo

CURSO ACADÉMICO: 2023-2024

Universida_{de}Vigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE MÁSTER

*Gemelo Digital de Entorno Operativo Marítimo: Propuesta de
arquitectura de integración de datos y operación*

Máster Universitario en Dirección TIC para la Defensa
Especialidad de Sistemas y Tecnologías de la Información

Universida_{de}Vigo

RESUMEN

El mundo experimenta la Cuarta Revolución Industrial, marcada por tecnologías disruptivas como la inteligencia artificial, *Big Data*, el internet de las cosas, etc., habilitadoras del denominado gemelo digital, el cual, basándose en la disponibilidad de datos masivos sobre un determinado proceso o sistema, crea un modelo virtual del mismo capaz de proporcionar valiosa información del funcionamiento actual y futuro. En este contexto, instituciones meteorológicas y marítimas han reconocido el valor de datos y modelos predictivos abiertos para mejorar, entre otras, la planificación y operación marítima. Esta información, junto con la que las propias plataformas marítimas son capaces de generar, debidamente integradas mediante técnicas de inteligencia artificial y *Big Data*, constituyen una base prometedora para el desarrollo de un gemelo digital de un entorno marítimo operativo que permita a las plataformas disponer de una mayor consciencia situacional en su entorno de operaciones, lo que se traducirá en una mayor seguridad, efectividad y eficiencia en el cumplimiento de su misión. Sin embargo, las limitaciones de capacidad informática y comunicaciones en plataformas navales plantean desafíos. Para abordarlos, se plantea el diseño conceptual de una arquitectura de integración de datos que permita el desarrollo de ese gemelo digital basada en tres niveles: un sistema central terrestre que crea y mantiene el gemelo digital usando todas las fuentes disponibles; una plataforma naval que ejecuta un modelo local actualizado con datos propios, y plataformas desplegables que proporcionan información adicional. Asociadas a esta arquitectura se definen las estrategias de operación que permiten su funcionamiento.

PALABRAS CLAVE

Gemelo digital oceánico, entorno operativo marítimo, operaciones marítimas, GIS

AGRADECIMIENTOS

A Cecilia por su ánimo, apoyo y toneladas de paciencia.

A mis tutores, Milagros y Carlos, por su labor y ayuda.

Contenido

Índice de Figuras	3
Índice de Tablas.....	5
Abreviaturas	6
1 Introducción y objetivos	9
1.1 Objetivos	9
1.2 Estructura del documento.....	10
2 Estado del arte	12
2.1 Gemelo digital	12
2.1.1 Qué es un gemelo digital	12
2.1.2 Tipos de gemelos digitales.....	16
2.2 Gemelo digital de entorno marítimo y proyectos relacionados	17
2.2.1 European Digital Twin of the Ocean (EDITO)	17
2.2.2 Ocean Twin.....	19
2.2.3 Programa DECADE – DITTO.....	20
2.2.4 Gemelo digital del entorno operativo marítimo.....	21
2.3 Formatos de datos oceanográficos y fuentes de datos institucionales y propias	21
2.3.1 Formatos de datos oceanográficos	21
2.3.2 Fuentes de datos institucionales.....	26
2.3.3 Fuentes de datos propias de la plataforma desplegada	38
2.4 Gestión y almacenamiento de los datos	42
2.4.1 Datos estructurados vs no Estructurados	43
2.4.2 Bases de datos SQL y NoSQL	43
2.4.3 Big Data	47
2.5 Sistemas de comunicaciones en el entorno naval	53
2.6 Sistemas de cuadrícula jerárquica para modelado del entorno	56
2.6.1 Geohash	57
2.6.2 S2	57
2.6.3 Teselas de Bing.....	58
2.6.4 H3 de Uber.....	59
3 Arquitectura del sistema	60
3.1 Arquitectura global.....	60
3.2 Nivel 1: Servidor central en tierra.....	61
3.2.1 Información proporcionada por el gemelo digital	65
3.2.2 Interfaces del componente	67

3.3 Nivel 2: Plataforma naval con capacidades medias de procesamiento de datos.....	68
3.3.1 Información proporcionada por el Nivel 2	69
3.3.2 Interfaces del componente	70
3.4 Nivel 3: plataformas auxiliares desplegadas desde la plataforma principal	71
3.4.1 Información proporcionada por el Nivel 3	71
3.4.2 Interfaces del componente	72
4 Operación del Sistema.....	73
4.1 Operación del gemelo digital: casos de uso de integración	73
4.1.1 Escenario 1: Sin restricciones de conectividad entre Nivel 1 y Nivel 2	74
4.1.2 Escenario 2: Conectividad restringida entre Nivel 1 y Nivel 2.	77
4.1.3 Escenario 3: Sin conectividad en la plataforma de Nivel 2 con el Nivel 1.....	79
4.2 Discusión de la solución	80
5 Conclusiones y líneas futuras	82
Bibliografía.....	84

ÍNDICE DE FIGURAS

Figura 2-1: Modelos simplificados de control de los tipos de gemelos digitales [10].....	17
Figura 2-2: Modelo conceptual de <i>EDITO</i> [11].....	18
Figura 2-3: Diseño de una arquitectura del gemelo digital del océano [11]	19
Figura 2-4: Catálogo de fuentes ODIS [24]	27
Figura 2-5: Catálogo de datos de Copernicus [3]	32
Figura 2-6: Ejemplo de comando de descarga de datos a través de <i>API</i> [3].....	33
Figura 2-7: Página de Puertos del Estado para visualización y descarga de datos [31]	35
Figura 2-8: Medidores de velocidad y dirección del viento. A la Izquierda uno de la empresa Campbell Scientific y a la derecha uno más sofisticado de la empresa Vaisala [35] y [36]	39
Figura 2-9: Sensor de temperatura y humedad relativa HygroVUE10 de Campbell Scientific [36].....	39
Figura 2-10: Medidor de visibilidad de Campbell Scientific [36]	40
Figura 2-11: Medidor de tiempo actual [35]	40
Figura 2-12: Vaisala Ceilometer CL31 [35]	40
Figura 2-13: Detector de relámpagos BTM-350 de Observator Instruments [37]	41
Figura 2-14: Comparativa de <i>BBDD SQL</i> y <i>NoSQL</i> [46]	44
Figura 2-15: Comparativa de características de <i>BBDD SQL</i> y <i>NoSQL</i> [47].....	44
Figura 2-16: Representación gráfica del teorema CAP [49]	46
Figura 2-17: <i>Flujo de Analítica de Datos</i> [51].....	48
Figura 2-18: Herramientas destacadas para implementación de <i>Big Data</i> [52]	50
Figura 2-19: Esquema de <i>Data warehouse</i> . Fuente https://corporatefinanceinstitute.com/	51
Figura 2-20: Esquema de <i>Data lake</i> . Imagen de datakitchen.io	52
Figura 2-21: Esquema de un Data Lakehouse [54].....	52
Figura 2-22: Imagen promocional del alcance de las soluciones <i>satcom</i> navales de <i>Thales</i> [56]	53
Figura 2-23: Representación gráfica de la arquitectura <i>IP</i> de comunicaciones navales de <i>Thales</i> [57].	54
Figura 2-24: Diferentes soluciones de comunicaciones navales de Leonardo [58].....	55
Figura 2-25: Diagrama de bloques de los sistemas de comunicaciones navales de <i>Leonardo</i> [58]	56
Figura 2-26: Ejemplo de cuadrícula <i>GeoHash</i> [59]	57
Figura 2-27: Ejemplo de cuadrícula <i>S2</i> [549]	58
Figura 2-28: Ejemplo de teselas <i>Bing</i> [59].....	58
Figura 2-29: Ejemplo de representación <i>H3</i> de parte de California [5560].....	59
Figura 3-1: Componentes del sistema del Gemelo Digital de Entorno Operativo Marítimo [Elaboración propia con imágenes obtenidas de depositphotos.com]	61
Figura 3-2: Arquitectura del sistema central (Nivel 1)	65

Figura 3-3: Ejemplo de representación de variables del entorno marítimo usando el modelo de celdas <i>H3</i> de <i>Uber</i> [61]	66
Figura 3-4: Arquitectura de integración de datos de la plataforma naval (Nivel 2)	69
Figura 4-1: Caso de uso de actualización del modelo de gemelo digital para su ejecución en la plataforma (Escenario 1)	75
Figura 4-2: Caso de uso de descarga de datos del entorno operativo de la misión a la plataforma (Escenario 1)	76
Figura 4-3: Caso de uso de descarga de datos del entorno operativo al sistema central (Escenario 1)	77
Figura 4-4: Caso de uso de decarga de datos del entorno operativo de la misión a la plataforma (Escenario 2)	78
Figura 4-5: Caso de uso de envío de informes de las plataformas desplegadas al sistema central (Escenario 2)	79
Figura 4-6: Caso de uso de funcionamiento de la plataforma en modo stand-alone (Escenario 3)	80

ÍNDICE DE TABLAS

Tabla 2-1: Desafíos comunes y tecnologías habilitadoras del gemelo digital [6].....	13
Tabla 2-2: Tipos de Gemelos Digitales [10]	16
Tabla 2-3: Ejemplo de fichero <i>JSON</i> [18].....	23
Tabla 2-4: Ejemplo de ficheros <i>GeoJSON</i> y <i>TopoJSON</i> [15].....	24
Tabla 2-5: <i>EMODNet</i> : Solicitud de metadatos de servicio [27].....	28
Tabla 2-6: <i>EMODNet</i> : Solicitud de registros [27].....	28
Tabla 2-7: <i>EMODNet</i> : Solicitud de registros mediante identificador [27].....	29
Tabla 2-8: <i>EMODNet</i> : Servicios <i>WFS</i> [27]	30
Tabla 2-9: <i>EMODNet</i> : Servicios <i>WCS</i> [27].....	31
Tabla 2-10: <i>EMODNet</i> : Servicios <i>No OGC</i> [27].....	31
Tabla 2-11: Acceso automatizado a datos de Compernicus [3]	34
Tabla 2-12: Comparativa de <i>BBDD</i> Relacionales y <i>NoSQL</i> [48].....	45
Tabla 2-13: Principales características de soluciones conocidas de <i>BBDD NoSQL</i> [47]	47

ABREVIATURAS

Abreviatura	Definición
2D	2 dimensiones
3D	3 dimensiones
ACSP	Acoustic Doppler Current Profiler
AEMET	Agencia Estatal de Meteorología
AHRPT	Transmisión Avanzada de Imágenes de Alta Velocidad
API	Application Programming Interface
AR	Augmented Reality
AWS	Amazon Web Services
BBDD	Base/s de Datos
BI	Business Intelligence
BUFR	Formato Universal Binario para Representación de Datos Meteorológicos
CAP	Protocolo de Alerta Común
CFMS	The Coordination Group for Meteorological Satellites
CLI	Command Line Interface
CNN	Convolutional Neural Network
CPD	Centro de Procesamiento de Datos
CPU	Central Processing Unit
CRM	Customer Relationship Management
CSV	Comma Separated Values
CSW	Catalog Service for the Web
DaaS	Documents as a Service
DestineE	Destino Tierra
DITTO	Digital Twin of the Ocean
DL	Deep Learning
DTED	Digital Terrain Elevation Data
DTM	Digital Terrain Model
DTO	Digital Twin of the Ocean
DVL	Doppler Velocity Log
DW	Data Warehouse
EDITO	European Digital Twin of the Ocean
EMODNET	European Marine Observation and Data Network
EO	Electrónico
EOP	Earth Observation Product
EOS	Earth Observing System
ERDDAP	Environmental Research Division's Data Access Program
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
FAIR	Localizable, accesible, interoperable y reusable
GAN	Generative Adversarial Network
GDAL	Geospatial Data Abstraction Library
GEBCO	The General Bathymetric Chart of the Oceans
GFS	Global Forecast System

Abreviatura	Definición
GIS	Geographic Information System
GML	Geography Markup Language
GPU	Graphics Processing Unit
GRIB	Formato Binario Estándar de la WMO para Datos en rejilla
HDF	Hierarchical Data Format
HF	High Frequency
HPC	High Power Computing
HRIT	Transmisión de Información de Alta Velocidad
HTTP	Hypertext Transfer Protocol
IA	Inteligencia Artificial
IEO	Instituto Español de Oceanografía
IOC	Intergovernmental Oceanographic Commission
IODE	Intercambio Internacional de Datos e Información Oceanográfico
IoT	Internet de las cosas
IP	Internet Protocol
IR	Infrarrojo
JPEG	Joint Photographic Experts Group
JSON	JavaScript Object Notation
Kbps	Kilo bits per second
KML	Keyhole Markup Language
KVP	Key-Value Pair
Mbps	Mega bits per second
McIDAS	Sistema Interactivo de Acceso a Datos por Computadora
MDA	Conciencia del Dominio Marítimo
MF	Medium Frequency
MHS	Message Handling System
ML	Machine Learning
MOHID	Modelo Hidrodinámico
MR	Mixed Reality
NCEP	National Energy and Climate plans
NetCDF	Network Common Data Form
NITFS	National Imagery Transmission Format Standard
NMEA	Asociación Nacional de Electrónica Marina
NoSQL	Not-only SQL
OASIS	Organización para el Avance de los Estándares de Información Estructurada
ODIS	Oceanic Data and Information System
OGC	Open Geospatial Consortium
OLCI	Ocean and Land Color Instrument
OPeNDAP	Open-source Project for a Network Data Access Protocol
PCA	Análisis de Componentes Principales
PNG	Portable Network Graphic
REST	REpresentational State Transfer
RHIB	rigid-hulled inflatable boat
ROMS	Regional Ocean Modeling System
RPAS	Remotely Piloted Aircraft System

Abreviatura	Definición
RPC	Remote Procedure Call
SDR	Software Defined Radio
SFTP	Secure File Transfer Protocol
SIG	Sistema de Información geográfica
SIP	Submission Information Package
SLSTR	Radiómetro de temperatura de la superficie del mar y de la tierra
SOAP	Simple Object Access Protocol
SQL	Standard Query Language
SVM	Support Vector Machines
SWAN	Simulating Waves Nearshore
SWAT	Herramienta para la Evaluación del Suelo y Agua
TB	Terabyte
TFM	Trabajo Fin de Máster
THREDDS	Thematic Realtime Environmental Distributed Data Service
TIC	Tecnologías de la Información y Comunicaciones
TIFF	Tag Image File Format,
TPU	Tensor Processing Unit)
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
UE	Unión Europea
UHF	Ultra High Frequency
UML	Unified Modeling Language
URL	Uniform Resource Locator
USV	Unmanned Surface Vehicle
UTC	Universal Time Coordinated
UUV	Unmanned Underwater Vehicle
VHF	Very High Frequency
VoIP	Voice over IP
VR	Virtual Reality
WCS	Web Coverage Service
WFS	Web Feature Service
WMO	Organización Meteorológica Mundial
WMS	Web Map Service
WMTS	Web Map Tile Service
WRF	Weather Research Forecast
WW3	Wave Watch III
XML	eXtensible Markup Language
ZIP	Formato de archivo comprimido

1 INTRODUCCIÓN Y OBJETIVOS

El mundo está inmerso en la que se denomina Cuarta Revolución Industrial o Industria 4.0 que, al igual que las anteriores, implica un importante cambio de paradigma en las organizaciones a todos los niveles, tanto estratégico como organizativo y operativo [1]. Esta revolución viene propiciada, principalmente, por la aparición de nuevas tecnologías que se han denominado disruptivas, como la inteligencia artificial (IA), el *Big Data*, la robótica, el internet de las cosas (IoT), las comunicaciones 5G, el *blockchain*, etc [2]. De forma más o menos directa, todas estas tecnologías tienen un elemento común que son los datos. Estos se generan, procesan, transmiten y consumen para aprovechar todo su potencial y convertirlos en información útil que permita a las organizaciones mejorar sus procesos, ser más eficaces y eficientes, tomar mejores decisiones y así mejorar su posicionamiento y adaptarse más fácilmente en un mundo cada vez más complejo y dinámico. Es el tiempo de las organizaciones centradas y dirigidas por los datos.

Dentro de esta revolución basada en los datos, las instituciones de meteorología, hidrográficas, oceanográficas, etc., también se han dado cuenta del valor que estos activos pueden proporcionar a la sociedad. En los últimos años, algunas de estas organizaciones están desplegando plataformas donde publican de forma abierta una gran cantidad de datos e información de condiciones meteorológicas y marítimas, tanto valores actuales como predicciones y actualizados de forma periódica cada pocas horas [3]. Esta información puede ser de gran utilidad para multitud de fines, entre ellos, para la planificación y operación en las misiones marítimas.

Por otro lado, las propias plataformas navales dotadas de múltiples sistemas de sensores, especialmente las plataformas de Defensa, también generan una gran cantidad de datos en tiempo real del entorno en el que operan [4]. Estos datos también podrán ser explotados para mejorar la consciencia del entorno y los modelos predictivos, reducir la incertidumbre y tomar mejores decisiones que ayuden a navegar de forma más segura, eficaz y eficiente.

Con tal cantidad de datos disponibles de diferentes fuentes y las nuevas tecnologías mencionadas, se plantea la idea de desarrollar un sistema basado en *Big Data* e IA que permita integrarlos para crear y actualizar unos modelos predictivos que puedan ser utilizados sobre el terreno por las plataformas, constituyendo lo que ahora se denomina **gemelo digital** [5], en este caso de las condiciones marítimas y climatológicas para la navegación. Sin embargo, la potencia requerida para desarrollar, actualizar y mantener este modelo requiere grandes CPDs (Centros de Proceso de Datos) imposibles de instalar en plataformas navales, las cuales podrán, en todo caso, hacer uso de dichos modelos alimentándolos con los datos obtenidos localmente y dejando la parte de computación intensiva a sistemas terrestres. Pero, además, este escenario naval presenta otra complicación referente a las comunicaciones ya que, cuando se está lejos de la costa, las plataformas dependen de las comunicaciones satelitales o radio de largo alcance, que ofrecen anchos de banda limitados insuficientes para una realimentación continua de los nuevos datos al gemelo digital.

Abordar estas problemáticas, requiere analizarlas detenidamente y plantear alguna solución que permita explotar toda esa información disponible, siempre con la mira puesta en la mejora de la operatividad de las misiones marítimas. Esto va a requerir el diseño de una arquitectura de integración de la información entre diferentes componentes del sistema y cuyo funcionamiento va a implicar el establecimiento de una estrategia de operación.

1.1 Objetivos

El objetivo de este Trabajo de Fin de Máster (TFM) es el diseño de una arquitectura conceptual de integración de datos que habilite el desarrollo, explotación y actualización de un gemelo digital de un

entorno operativo marítimo. La arquitectura propuesta aborda las limitaciones en cuanto a capacidad informática instalable en plataformas navales y las limitaciones en las comunicaciones operativas que dificultan la realimentación de los datos al gemelo digital. Para ello, se propone una arquitectura por niveles, donde existe un primer nivel donde un sistema central terrestre, con gran potencia de cómputo y comunicaciones, que crea y mantiene el gemelo digital usando todas las fuentes de datos disponibles. En el segundo nivel la plataforma naval, que desplegará en escenarios operativos, con comunicaciones limitadas, ejecutará un modelo local del gemelo digital actualizado con datos propios y datos proporcionados por el sistema central bajo petición. En el último nivel se situarían las plataformas desplegables desde el segundo nivel, que proporcionarán información adicional. Para optimizar la integración de la información y la operatividad de las plataformas en estos contextos operativos con limitaciones, se plantean una serie de estrategias de operación y se analizan los pros y contras de la solución propuesta. El modelo conceptual resultante podrá servir de marco de referencia para una futura implementación real.

Para alcanzar el objetivo, en primer lugar, se introducirá una base teórica de lo que es un gemelo digital y, en concreto, un gemelo digital marítimo, identificando requisitos, restricciones, aplicaciones y proyectos existentes, con especial atención a las fuentes de datos disponibles, formatos, almacenamiento, gestión, intercambio y explotación de los mismos. Se presentará un pequeño estudio del estado del arte de los principales conceptos y tecnologías implicadas.

Tomando como base dicho estudio, se propondrá la arquitectura que se adapte al escenario mencionado, así como las estrategias de operación asociadas, tales que permitan un aprovechamiento óptimo de toda la información disponible y que se traduzca en la mejora operativa de las misiones marítimas.

Se quiere recalcar que el enfoque de este trabajo no es de carácter profundamente técnico, sino de tipo conceptual, enfocándose más en las ideas y posibilidades y no tanto en las implementaciones. No obstante, se han utilizado multitud de referencias que aparecen indicadas a lo largo del texto y que podrán consultarse para profundizar en cada una de las materias.

1.2 Estructura del documento

El TFM se ha estructurado de la siguiente manera:

- En el capítulo 2 se presentan, en primer lugar, los fundamentos teóricos de los gemelos digitales, el estado actual y aplicaciones de este tipo de tecnologías, así como su uso en el contexto oceánico incluyendo los principales proyectos en los que se está trabajando. Seguidamente, se abordan los principales formatos de datos, así como las fuentes de datos disponibles, tanto a nivel institucional de fuentes abiertas o propietarias, como a nivel de plataformas. Teniendo en cuenta estas fuentes de datos, se identifican y describen las opciones tecnológicas actuales para el almacenamiento, gestión y explotación de los datos que se pueden adaptar a los requisitos. Después, se presenta un breve estudio de los sistemas de comunicación navales actuales y sus características, especialmente en lo referente a arquitectura, capacidad, disponibilidad, seguridad y operatividad. Por último, se hace una pequeña introducción a los sistemas de representación *GIS* (*Geographic Information System*) basados en rejillas o cuadrículas, por considerarse la mejor opción para integrar los datos en el gemelo digital de entorno operativo.
- En el capítulo 3, se propone una arquitectura a nivel conceptual del sistema, describiendo los componentes principales del sistema y los diferentes bloques funcionales.

- En el capítulo 4, teniendo en cuenta la arquitectura y las limitaciones operativas, especialmente en lo referente a comunicaciones, se establecerán diferentes posibilidades y estrategias de operación para el intercambio de datos e información. Estas se acompañarán de una serie de diagramas de casos de uso que permitan ver de forma más gráfica como se abordan los diferentes escenarios de funcionamiento. También se incluirá un análisis de pros y contras de la solución propuesta.
- El capítulo 5 incluirá las conclusiones sobre la problemática planteada y las soluciones adoptadas, así como las posibles líneas de trabajo futuras.
- Finalmente se presentarán las referencias bibliográficas consultadas para el desarrollo del *TFM*.

2 ESTADO DEL ARTE

En este capítulo se pretende dar una visión general del estado del arte de todos aquellos campos de trabajo, tecnologías o elementos que forman parte o pueden contribuir al desarrollo del gemelo digital oceánico para un entorno operativo. Hay que tener en cuenta que un gemelo digital de algo tan complejo como es el océano y todas las variables involucradas requiere un enfoque multidisciplinar cuyo alcance es inmenso e inabordable en un trabajo como el presente. Por tanto, se han seleccionado los elementos más significativos relacionados con las fuentes de datos y la arquitectura de almacenamiento, gestión e intercambio de los mismos para alimentar y utilizar el gemelo digital en el entorno operativo de una plataforma naval de Defensa.

2.1 Gemelo digital

Esta sección pretende dar unas nociones básicas de lo que se esconde detrás del concepto “Gemelo Digital” del que se viene hablando mucho en los últimos años; cuáles son sus objetivos y ventajas; y, puesto que este capítulo trata sobre el estado del arte, se quieren recalcar las complejas problemáticas implicadas en su implementación y las tecnologías aplicadas actualmente para ello.

2.1.1 Qué es un gemelo digital

Adil Rasheed, et al [6] definen el gemelo digital como “una representación virtual de un activo físico habilitada a través de datos y simuladores para predicción, optimización, seguimiento, control y mejora en la toma de decisiones en tiempo real”. No es casual que en los últimos años se haya convertido el gemelo digital en una tendencia importante y emergente en muchas aplicaciones, ya que es gracias a los recientes avances en sistemas informáticos, modelado de sistemas físicos, inteligencia artificial, *Big Data*, cibernética, el procesamiento de datos y las herramientas de gestión, lo que permiten que los gemelos digitales puedan ser una realidad.

Además, la misma referencia anterior nos indica algunas de las características del gemelo digital que aportan especial valor, entre las que se destacan las siguientes:

- 1) **Monitorización y control remoto en tiempo real:** normalmente, es prácticamente imposible tener un acceso profundo a un sistema físico grande en tiempo real. Un gemelo digital será accesible en cualquier lugar no solo para monitorizar su rendimiento, sino también controlarlo de forma remota.
- 2) **Mayor eficiencia y seguridad,** al permitir reducir la necesidad de intervención humana en muchos trabajos peligrosos o repetitivos que pueden ser asignados a robots, con supervisión humana remota cuando sea necesario.
- 3) **Mantenimiento predictivo y programación:** el uso de múltiples sensores que proporcionan grandes cantidades de datos permitirá realizar análisis y detección de fallos con gran antelación permitiendo desarrollar y optimizar los programas de mantenimiento, incluyendo preventivos y prescriptivos.
- 4) **Escenario y evaluación de riesgos:** mediante un modelo digital fiel al real se podrán ejecutar análisis de hipótesis creando escenarios sintéticos que permitirán una evaluación de riesgos, así como desarrollar estrategias de mitigación, todo ello sin poner en riesgo el activo real.
- 5) **Apoyo en la toma de decisiones:** una mayor disponibilidad de datos, unida a análisis avanzados en tiempo real, ayudará en la toma de decisiones con más información y más rápidamente.

- 6) **Personalización de productos y servicios:** basándose en requisitos históricos, preferencias actuales y la evolución del mercado, el gemelo digital permitirá una adaptación más rápida y fluida para atender las necesidades cambiantes.
- 7) **Mejor documentación y comunicación:** la información disponible y accesible en tiempo real combinada con informes automatizados ayudarán a mantener bien informadas a las partes interesadas, mejorando así la transparencia.

En la siguiente tabla se resumen los desafíos que implica el desarrollo de los gemelos digitales y las tecnologías que posibilitan abordar estos desafíos.

Desafíos	Tecnologías habilitadoras
Gestión de los datos, privacidad, seguridad y calidad de los datos	Plataformas digitales Tecnologías criptográficas y blockchain Tecnologías Big Data
Comunicación de datos en tiempo real y latencia	Compresión de los datos Tecnologías de comunicaciones como 5G Tecnologías de IoT
Realismo en el modelado físico y proyecciones futuras	Tecnologías de sensores Simuladores de alta fidelidad basados en la física de los procesos Modelos data-driven
Modelado en tiempo real	Análisis y modelado híbrido Modelado de orden reducido Modelos multi-variable data-driven
Actualización continua del modelo, modelado de lo desconocido	Big Data y cibernética Modelado y análisis multivariable Asimilación de datos ¹ “Sensado” comprimido y regresión simbólica
Transparencia e interoperabilidad	Análisis y modelado híbrido Inteligencia artificial explicable
Computación a gran escala	Infraestructura computacional, computación edge, fog y cloud.
Interacción con el activo físico	Interfaz hombre-máquina Procesado del lenguaje natural Realidad aumentada y realidad virtual.

Tabla 2-1: Desafíos comunes y tecnologías habilitadoras del gemelo digital [6]

Como se podrá ir viendo a lo largo del trabajo, la problemática abordada por este se enfrenta a prácticamente todos los desafíos mencionados, con unas complejidades específicas añadidas provenientes del tipo de plataformas de tipo naval y su correspondiente entorno operativo. No obstante,

¹La asimilación de datos combina observaciones y modelos de una manera que tenga en cuenta las incertidumbres de cada uno, respetando al mismo tiempo ciertas limitaciones [7].

dentro del alcance del trabajo solo se abordarán algunos de los desafíos mencionados y las tecnologías habilitadoras más relevantes, centrándose en la arquitectura de sistemas y en el procesado y gestión de los datos, estrategia de comunicaciones y actualización de los modelos.

2.1.1.1 Modelado basado en datos (data-driven)

Aunque este trabajo no está enfocado en la parte de modelado del sistema, sí que se asume que el sistema que conforma el gemelo digital completo se basará principalmente en un modelado del tipo *data-driven*, que, como su propio nombre indica, se basa en los datos, cuya obtención, almacenamiento, gestión e intercambio sí son el foco del trabajo.

Un modelo *data-driven* consiste básicamente en el desarrollo de un modelo complejo utilizando técnicas de aprendizaje automático o *Machine Learning (ML)* sobre los datos obtenidos, tanto a partir de la operación del sistema real utilizando para ello sensores de diferentes tipos, como de fuentes externas fiables. Los modelos *data-driven* son cada vez más populares debido a la cada vez mayor disponibilidad de datos a partir de sensores baratos y miniaturizados de todo tipo; existencia de potentes librerías de procesamiento y aprendizaje (*openAI*, *tensorflow*, etc.); el acceso asequible a potente infraestructura computacional (*CPU*, *GPU* y *TPU*), así como a recursos formativos.

Una de las principales ventajas de este tipo de modelos es su fidelidad, ya que, por su naturaleza, cubren la física completa de los sistemas, tanto la conocida como la no conocida. Por contra, algunas técnicas de *Machine Learning* más usadas como el aprendizaje profundo proporcionan modelos tipo caja negra, o “no explicables”.

A continuación, se resume brevemente el estado del arte del modelado basado en datos para la implementación de gemelos digitales, considerando aspectos que van desde la generación y seguridad de datos hasta el uso de técnicas avanzadas [6].

- 1) **Preprocesamiento, Gestión y Propiedad de Datos:** Con el aumento de la cantidad de datos, surgen desafíos relacionados con la calidad, gestión y titularidad de la información. Es imperativo mejorar la calidad de los datos y comprimirlos en tiempo real para garantizar un rendimiento eficiente del gemelo digital. Existen algoritmos como Componentes Principales (*PCA*) y técnicas más avanzadas, incluyendo máquinas de *Boltzman* restringidas [8], para abordar problemas como la detección automática de valores atípicos y el rellenado de datos faltantes. También se están explorando métodos innovadores para el aprendizaje continuo a partir de flujos de *Big Data*. Por otro lado, la propiedad de los datos emerge como un desafío significativo, ya que involucra a diversos actores en la generación, procesamiento y análisis de datos, lo cual podría comprometer el funcionamiento efectivo del gemelo digital.
- 2) **Privacidad de Datos y Cuestiones Éticas:** La implementación de gemelos digitales útiles requiere un alto grado de confianza en el sistema de información, especialmente en un entorno donde las leyes sobre privacidad son cada vez más estrictas. Surge el concepto de *blockchain* como una solución revolucionaria, proporcionando un registro distribuido, seguro y verificable. Este enfoque garantiza la seguridad y trazabilidad de la información en gemelos digitales, abordando así preocupaciones clave en términos de privacidad y ética.
- 3) **Aprendizaje Automático e Inteligencia Artificial:** El análisis inteligente de datos mediante técnicas de *Machine Learning (ML)*, *Deep Learning (DL)* y otras técnicas de Inteligencia Artificial (*IA*) desempeñan un papel fundamental en el contexto de los gemelos digitales. Existen diversos enfoques, desde el aprendizaje supervisado, con algoritmos como regresión logística y máquinas de vectores de soporte (*SVM*), hasta técnicas más avanzadas, como las máquinas de *Boltzmann* restringidas, las Redes Neuronales Convolucionales (*CNN*) y las Redes Generativas Adversarias (*GAN*) [8]. Además, también se destaca la utilidad del aprendizaje por refuerzo en situaciones de falta de datos, por su potencial en la toma de decisiones inteligentes para la Industria 4.0.

Ventajas y Desventajas de Modelos Basados en Datos

Los modelos basados en datos (*data-driven*) presentan la ventaja de mejorar con la alimentación continua de datos, aunque la fase de entrenamiento puede experimentar inestabilidades. Sin embargo, una vez entrenados, estos modelos son estables para realizar predicciones. Por otro lado, modelos más complejos, especialmente aquellos basados en Redes Neuronales Profundas (*DNN*), pueden carecer de interpretabilidad [9]. Este aspecto es crucial en aplicaciones críticas para la seguridad. Además, hay que tener en cuenta que los modelos pueden ser sesgados según los datos de entrenamiento.

Se plantean, por tanto, desafíos y enfoques en el modelado basado en datos para gemelos digitales, con aspectos cruciales como la calidad y propiedad de los datos, la privacidad, la ética y la aplicación de diversas técnicas de aprendizaje automático e inteligencia artificial.

2.1.1.2 Infraestructuras tecnológicas

La implementación efectiva de modelos de aprendizaje automático, *Big Data* y otras tecnologías innovadoras habilitadoras del gemelo digital depende en gran medida de las infraestructuras tecnológicas disponibles, destacando el *Big Data*, Internet de las Cosas (*IoT*), comunicaciones, infraestructuras computacionales y plataformas gemelas digitales que se describen brevemente a continuación:

- **Tecnologías de *Big Data*:** La gestión de grandes volúmenes de datos ha experimentado avances significativos. Las plataformas actuales ofrecen capacidades integrales para la manipulación, integración, almacenamiento y análisis de datos a gran escala permitiendo ejecutar tareas directamente sobre su almacenamiento, evitando la copia de datos a memoria local. La gestión eficiente del *Big Data* aborda no solo los avances tecnológicos, sino también algoritmos, aplicaciones y desafíos asociados. Posteriormente se abordará en profundidad esta cuestión.
- **Tecnologías de la *IoT*:** El auge de la *IoT* se ha convertido en una fuerza impulsora en diversos sectores, desde la atención médica hasta la agricultura. La disponibilidad de conexiones de alta velocidad, tecnología 5G, digitalización de sistemas y sensores de bajo costo ha contribuido a la expansión de la *IoT*. Plataformas líderes, como *Amazon Web Services (AWS)* y *Azure IoT Suite*, ofrecen soluciones confiables para diversas aplicaciones, desde investigación biomédica hasta tecnologías para hogares inteligentes.
- **Tecnologías de comunicaciones:** La rápida comunicación de datos es crucial para el funcionamiento efectivo de los gemelos digitales. La tecnología 5G, con su baja latencia y velocidad sin precedentes, se destaca como fundamental. Además, tecnologías como *LoRaWAN* y la colaboración con sistemas aéreos no tripulados (*UAS*) se están explorando para mejorar la conectividad en entornos concurridos o emergencias.
- **Infraestructuras computacionales:** el aumento de la potencia de los dispositivos con un coste reducido, unidos a la computación en la nube, la informática de borde (*edge computing*) y la computación de niebla (*fog computing*) ofrecen enfoques flexibles y eficaces para gestionar grandes cantidades de datos. La computación cuántica también se vislumbra como una opción prometedora para resolver problemas complejos.
- **Plataformas Gemelas Digitales:** Empresas como *Kongsberg Digital*, *MapleSoft*, *Cognite*, *Siemens*, *ANSYS* y *Oracle* lideran el desarrollo de plataformas gemelas digitales. Estas plataformas permiten la creación de modelos dinámicos precisos y la integración de datos en

tiempo real para mejorar el rendimiento, predecir fallos y optimizar el mantenimiento en diversas industrias.

2.1.2 Tipos de gemelos digitales

Cor Verdouw y otros [10] describen las diferentes fases del ciclo de vida de un gemelo digital que serían las de diseño, implementación, explotación y retiro. La fase más relevante será la de explotación en la que los gemelos digitales están conectados a su gemelo físico real, y durante la cual los gemelos digitales se utilizarán para monitorizar la situación del sistema real, estado de los objetos, prescribir estados deseados, predecir estados futuros y corregir de forma remota el estado de objetos del mundo real. En las fases previas, se pueden crear gemelos digitales para definir y simular los estados y el comportamiento de sus gemelos del mundo real. Por último, en la fase de retiro, los gemelos digitales pueden permanecer conceptualmente vivos y ser utilizados para recordar los estados históricos de objetos del mundo real. En base a lo anterior en el citado artículo se definen diferentes tipologías de gemelos digitales que se resumen en la *Tabla 2-2*:

Tipología	Definición	Uso
Gemelo Digital Imaginario	Entidad conceptual que representa un objeto que aún no existe en el mundo real. Se obtiene a partir de información sobre el gemelo físico, como requisitos funcionales, modelos 3D, especificaciones de materiales, modelos de producción, etc.	Puede simular el comportamiento de objetos diseñados que aún no existen dentro de las normas de tolerancia.
Gemelo Digital de Monitorización	Representación digital del estado, comportamiento y evolución de un objeto físico en tiempo (casi) real.	Monitoriza el estado, operación y entorno externo del objeto conectado. Puede ser descriptivo o de diagnóstico, proporcionando información sobre lo que sucede o explicando por qué sucede.
Gemelo Digital Predictivo	Proyección digital de estados y comportamientos futuros de objetos físicos utilizando análisis predictivos, simulación y aprendizaje automático	Realiza predicciones dinámicas basándose en datos (casi) en tiempo real del gemelo físico.
Gemelo Digital Prescriptivo	Objeto digital inteligente que recomienda acciones correctivas y preventivas sobre objetos reales.	Utiliza resultados de gemelos de monitorización y predictivos para sugerir acciones. Las decisiones son tomadas por humanos, pero las intervenciones pueden ser ejecutadas remotamente.
Gemelo Digital Autónomo	Opera de forma autónoma y controla completamente el comportamiento de objetos del mundo real sin intervención humana	Puede volverse auto-adaptativo, aprendiendo sobre su entorno, auto-diagnosticando necesidades de servicio y adaptándose a las preferencias de los usuarios
Gemelo Digital de Recolección	Mantiene el historial completo de un objeto físico que ya no existe en la vida real	Conforma la memoria digital de objetos pasados, siendo crucial para reducir impactos ambientales y optimizar objetos de próxima generación

Tabla 2-2: Tipos de Gemelos Digitales [10]

En la *Figura 2-1* se muestran los esquemas de control de las diferentes topologías de gemelos digitales mencionadas.

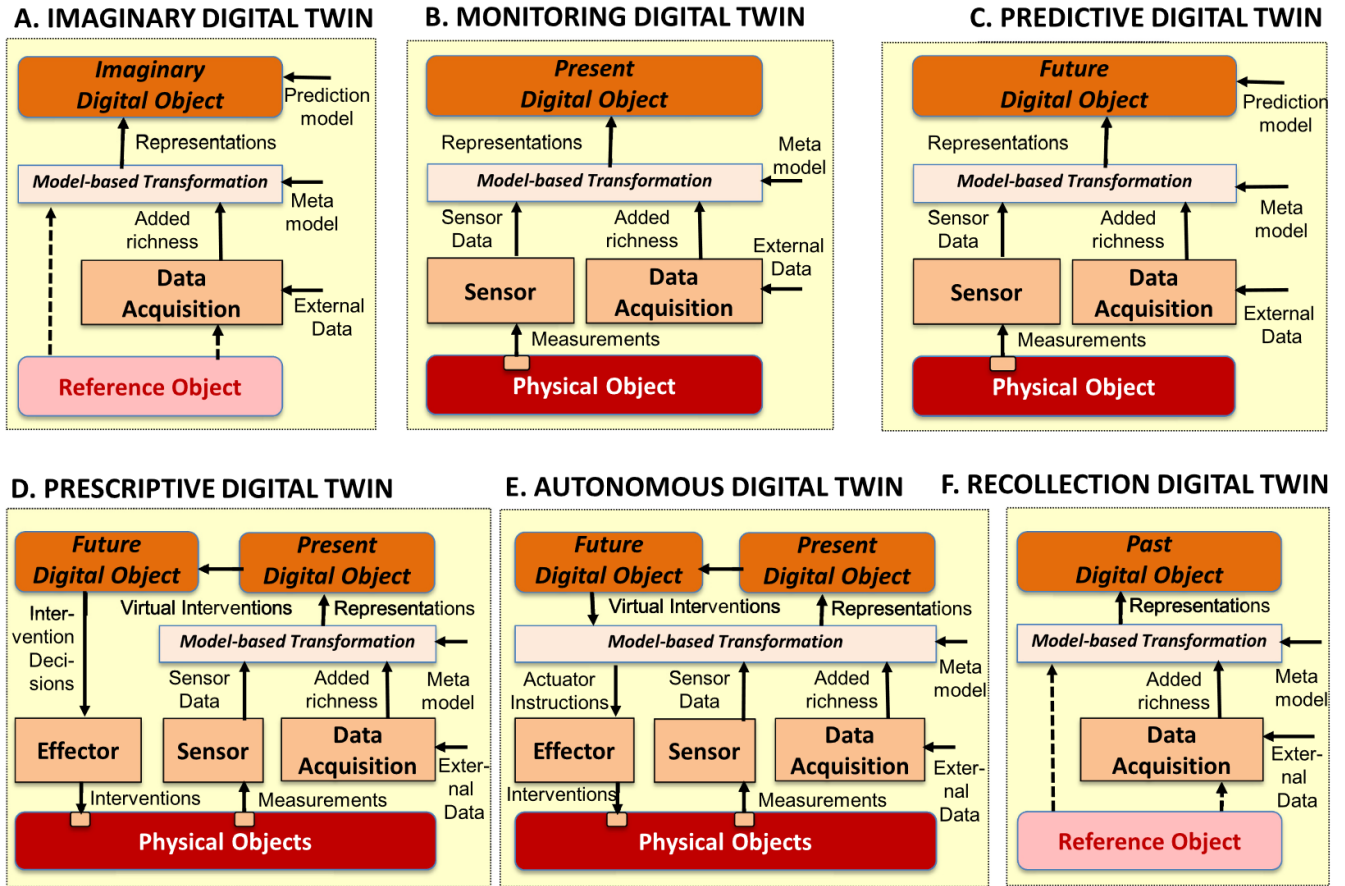


Figura 2-1: Modelos simplificados de control de los tipos de gemelos digitales [10]

2.2 Gemelo digital de entorno marítimo y proyectos relacionados

Una vez introducido el concepto general del gemelo digital, sus problemáticas, tecnologías implicadas y tipologías, se pretende en esta sección, por un lado, proporcionar una visión de lo que se entiende por gemelo digital oceánico a través de los diferentes proyectos existentes sobre esta materia. Por otro lado, una vez se tiene idea de lo que es un gemelo digital oceánico, se pretende concretar lo que se indica en el título de este trabajo como gemelo digital del entorno operativo marítimo.

A fecha de hoy, existen varios proyectos de alcance internacional en los que se está trabajando para la elaboración de gemelos digitales del océano (DTO). En las siguientes secciones se presentan algunos de los proyectos más significativos en este ámbito.

2.2.1 European Digital Twin of the Ocean (EDITO)

En la referencia [11] se define el gemelo digital oceánico como sigue:

“Un Gemelo Digital del Océano es una representación virtual de todos los entornos marinos y costeros del mundo. Da acceso a observaciones históricas y en tiempo real de miles de sensores en todo el océano mundial, así como a numerosos satélites en el espacio.

Al integrar modelos numéricos avanzados, inteligencia artificial, aprendizaje automático y computación de alto rendimiento, los gemelos digitales pueden generar la información necesaria para diseñar las formas más efectivas y sostenibles de proteger los hábitats marinos y costeros, apoyar una economía azul más sostenible y mitigar y adaptación al cambio climático.”

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99



It is important to note that the above results are based on the assumption that the data are stationary. If the data are non-stationary, the results may be biased. Therefore, it is important to test for stationarity before using the above methods.

Copernicus y contribuirá a consolidar *Destino Tierra (DestinE)*, una iniciativa de la Comisión Europea para desarrollar un modelo digital altamente preciso de la Tierra a escala mundial [12], garantizando el acceso a datos de diversas fuentes (*EMODnet*, *EuroGOOS*, *ECMWF*, *Copernicus Services* y observaciones del satélite *Sentinels*), instalaciones informáticas de alto rendimiento y conjuntos de modelos para diversas disciplinas oceánicas.

Este proyecto beneficiará a investigadores, expertos científicos, gobiernos, empresarios, y ciudadanos al poner los conocimientos sobre los océanos a su disposición. *EDITO-Model Lab* ofrecerá un Laboratorio Virtual de Modelos Oceánicos como una iniciativa colaborativa, proporcionando conjuntos de modelos, herramientas para usuarios intermedios y desarrolladores, con el objetivo de promover un océano seguro, saludable y productivo.

2.2.2 Ocean Twin

Como se puede ver en [13], el consorcio *Iliad*, bajo el nombre *INTEGRATED Digital Framework FOR Comprehensive MARITIME DATA AND INFORMATION SERVICES*, se embarca en el desarrollo de gemelos digitales del océano, aprovechando tecnologías como *IoT*, *IA*, *ML* y *Big Data*. El proyecto, financiado por la *UE*, tiene como objetivo crear representaciones virtuales del mar integrando observación de la Tierra, modelización y tecnologías digitales para prever futuros desarrollos en el entorno marino.

En la *Figura 2-3* se muestra una propuesta de un diseño conceptual de arquitectura de gemelo digital oceánico:



Figura 2-3: Diseño de una arquitectura del gemelo digital del océano [11]

2.2.2.1 Iliad DTO (Digital Twin of the Ocean): Gemelos Digitales Interoperables, Intensivos en Datos y Rentables

Iliad aspira a establecer gemelos digitales del océano interoperables, intensivos en datos y rentables. Capitalizando dos décadas de inversiones en políticas e infraestructuras para la economía azul, este proyecto fusionará datos de diversas fuentes de observación de la Tierra y avanzadas infraestructuras informáticas en un enfoque inclusivo y atractivo que contribuya a una economía oceánica sostenible.

2.2.2.2 Características Innovadoras de la Plataforma Iliad

La plataforma *Iliad* busca ser una red multiplataforma interoperable, integrada y rentable de sistemas de observación y predicción para los mares europeos, abarcando tanto el mar abierto como la zona costera. Sus características innovadoras incluyen la búsqueda unificada de datos, la descarga de información archivada/pronosticada a través de un único sistema, herramientas web interactivas y la disponibilidad de datos enriquecidos de diversas fuentes.

La plataforma reorganizará, homogeneizará y fusionará bases de datos para ofrecer datos en un formato estándar común. Además, se desarrollarán y probarán nuevos sistemas de observación, modelos operativos proporcionarán pronósticos, y se capacitará a usuarios finales y responsables de políticas locales/regionales/nacionales sobre el uso de los *DTO*.

2.2.2.3 Proyectos Piloto de Iliad

Iliad implementará proyectos piloto en áreas clave lideradas por socios industriales existentes. Estos incluyen energía eólica, energías marinas, pesca y acuicultura, tráfico marítimo y seguridad portuaria, contaminación, datos meteorológicos del océano y evaluaciones de biodiversidad. Estos pilotos utilizarán las tecnologías y habilidades existentes para recopilar y procesar datos, demostrando la aplicabilidad y eficacia de la plataforma *Iliad* en diversas aplicaciones marinas.

2.2.2.4 Iliad Marketplace: Un mercado para aplicaciones y servicios relacionados con geociencias

Para fomentar aplicaciones adicionales, *Iliad* creará el *Iliad Marketplace*, un mercado para aplicaciones y servicios relacionados con Geociencias. Similar a una tienda de aplicaciones, permitirá a los proveedores distribuir aplicaciones, *plug-ins*, interfaces y servicios derivados de gemelos digitales oceánicos. Esto proporcionará una forma eficiente para que los científicos descubran aplicaciones relevantes y servicios.

2.2.3 Programa DECADE – DITTO

El Decenio de las Ciencias Oceánicas para el Desarrollo Sostenible [14], lanzado por las Naciones Unidas en enero de 2021, tiene como objetivo impulsar una revolución en las ciencias oceánicas al fomentar la colaboración global y generar datos e información para respaldar políticas basadas en la ciencia. Durante esta década, se abordarán los 10 desafíos prioritarios, que incluyen la comprensión y superación de la contaminación marina, la protección de la biodiversidad, la alimentación sostenible de la población mundial y la creación de una economía oceánica sostenible. Para ello, la Alianza para el Decenio de los Océanos, compuesta por 10 patrocinadores y 18 miembros institucionales impulsa el respaldo al Decenio mediante la movilización de recursos, la creación de redes y la influencia en la comunidad global.

En consonancia con el Decenio, surge el programa *DITTO* (*Digital Twin of the Ocean*), establecido en 2021. *DITTO* se enfoca en desarrollar y compartir una comprensión común de los Gemelos Digitales del Océano (*DTO*), establecer mejores prácticas en su desarrollo y promover un marco digital para capacitar a profesionales oceánicos de todos los sectores del mundo.

DITTO respaldará el desarrollo de Gemelos Digitales del Océano como representaciones digitales realistas de activos, procesos o sistemas en entornos naturales o artificiales [15]. Esto se logrará mediante la creación de un marco digital que integre datos marinos, satelitales, modelado y simulación, computación de alto rendimiento, inteligencia artificial y herramientas especializadas. Este enfoque permitirá el acceso, manipulación, análisis y visualización efectivos de la información marina a nivel mundial.

Las actividades prioritarias de *DITTO* incluyen articular una comprensión común de los Gemelos Digitales del Océano, establecer procedimientos para generar confianza en ellos a través de acuerdos

justos y datos de calidad verificable, avanzar en prototipos de aplicaciones de Gemelos Digitales en diversos contextos y fomentar la formulación de proyectos relacionados con el Decenio.

2.2.4 Gemelo digital del entorno operativo marítimo

Como se ha podido leer en los proyectos de gemelos digitales del océano (DTO) que se han presentado, los objetivos son muy ambiciosos buscando desarrollar gemelos digitales del océano que abarcan multitud de disciplinas, desde las condiciones físicas de las masas de agua, hasta la parte biológica y efectos en el clima a nivel planetario.

Lo que todos tienen en común como elemento central del gemelo digital son los datos y su uso para la elaboración de modelos fidedignos útiles, esto es, la recopilación, almacenamiento, integración, análisis, intercambio, explotación y puesta a disposición de los usuarios de los datos, predicciones y modelos para todas aquellas partes interesadas.

Es debido a lo anterior que este trabajo tiene su razón de ser, ya que la disponibilidad de esos datos, predicciones y modelos pueden ser aprovechados e integrados en un sistema más específico cuyo objetivo sea el desarrollo y puesta a disposición de determinados grupos, de modelos de condiciones climatológicas y marítimas de lugares en los que se vayan a desarrollar operaciones marítimas. Pero este sistema no sólo aprovechará los datos provenientes de las fuentes y modelos externos, sino que las propias plataformas desplegadas producirán datos que se integrarán en el sistema para proporcionar información más exacta y actual de las zonas de operaciones que también contribuyan a la mejora y evolución de los modelos. De ahí la denominación de “gemelo digital en entorno operativo marítimo”, ya que el modelo del entorno local se realimenta a partir de los datos en vivo que proporcionan las plataformas, con el objetivo de reflejar de manera fiel las condiciones y producir predicciones precisas para la mejora de la conciencia situacional y con ello de la seguridad, eficacia y eficiencia de la misión.

Puesto que es en el capítulo 3 de este trabajo donde se abordará de manera detallada el sistema mencionado, no se darán en este punto más detalles al respecto. La idea de lo que se ha acaba de mencionar aquí es, básicamente, ayudar a contextualizar todo lo que se ha presentado y lo que sigue en el resto de este capítulo.

2.3 Formatos de datos oceanográficos y fuentes de datos institucionales y propias

En esta sección se presentarán en primer lugar algunos de los principales formatos de datos que se utilizan en el contexto oceánico y meteorológico. La segunda subsección proporciona información sobre algunas de las fuentes de datos institucionales más importantes disponibles para su utilización en el gemelo digital oceánico. Se incluyen fuentes de datos internacionales como *EMODNET*, o *Copernicus* y fuentes de datos nacionales como pueden ser *AEMET*, *Puertos del Estado* y *MeteoGalicia*. Por último, se indicarán algunas de las fuentes de datos propias de la plataforma sensorizada desplegada en un entorno operativo.

2.3.1 Formatos de datos oceanográficos

Como se verá en el siguiente apartado cuando se traten las fuentes de datos utilizables para alimentar el gemelo digital, existen numerosos formatos y servicios de acceso a datos que son de uso común en el contexto científico de observación de la Tierra, ya sean datos de estaciones de medición, generados sensores de todo tipo, o procedentes de imágenes satelitales. Por ello, se ha decidido proporcionar una descripción de los principales formatos y servicios de datos que aparecerán después.

Además, con esto también se quiere poner de relieve la variedad y heterogeneidad de los datos con el objetivo de, posteriormente, justificar por qué determinados modelos y arquitecturas de sistemas de almacenamiento son preferibles, o casi de obligado uso cuando se aborda un proyecto de esta magnitud y complejidad.

2.3.1.1 Estándares OGC (Open Geospatial Consortium) de datos GIS

OGC es una organización internacional sin fines lucrativos que agrupa miembros de organizaciones públicas y privadas y cuya razón de ser es la innovación, la colaboración y los estándares geoespaciales, principalmente a través de la creación de estándares abiertos e interoperables dentro del marco de los sistemas de información geográfica (GIS) e internet [16]. Apoya a empresas, agencias gubernamentales, organizaciones de investigación y universidades, todos trabajando conjuntamente para hacer que la información geoespacial sea *FAIR* (localizable, accesible, interoperable y reutilizable).

Los estándares de OGC se han convertido en la referencia mundial para el intercambio de datos GIS. Estos estándares definen tanto formatos de datos como servicios de acceso o intercambio. Entre los **formatos de datos** más extendidos que forman parte de los estándares OGC están [17]:

- **GML (Geography Markup Language):** Lenguaje de Marcado Geográfico basado en *XML (eXtensible Markup Language)* para representar información de elementos espaciales. Habilita el uso de un lenguaje común para el manejo e intercambio de información entre los diferentes Sistemas de Información Geográfica (SIG o GIS).
- **KML (Keyhole Markup Language):** desarrollado para *Google Earth*, desde 2008 KML es estándar OGC. También basado en *XML* permite representar datos geográficos en tres dimensiones. Un fichero KML especifica una característica (marcas de lugares, imágenes, polígonos, modelos 3D, descripciones textuales, etc.). Suelen distribuirse comprimidos como ficheros KMZ, que pueden incluir archivos de imágenes y otra información relacionada.

Respecto a los **estándares de servicios** de intercambio de datos geográficos a través de internet, OGC define los siguientes:

- **WMS (Web Map Service):** Define un interfaz *HTTP (Hypertext Transfer Protocol)* para solicitar imágenes de mapas georreferenciados. En la petición se indican las capas y el área de interés y se obtiene una o más imágenes (*JPEG, PNG, ...*). Permite además definir transparencia.
- **WFS (Web Feature Service):** A diferencia del anterior, que devuelve una imagen del mapa, WFS devuelve entidades con geometría y atributos que los clientes pueden utilizar en análisis geoespaciales, todo ello a través de *HTTP* en lugar de ficheros. Los servicios WFS también admiten filtros que permiten realizar consultas espaciales y de atributos en los datos. Los datos se transfieren en formato GML. El estándar define operaciones de descubrimiento, consulta, bloqueo y transacciones, para interrogar, descargar y editar características individuales.
- **WMTS (Web Map Tile Service):** Es un protocolo estándar para servir teselas de mapas georreferenciados por internet. Especifica peticiones de tipo *KVP (Key-Value Pair)*, *REST (REpresentational State Transfer)* y *SOAP (Simple Object Access Protocol)* a través de las cuales proporcionar teselas, información alfanumérica de para una localización en un mapa dada, y la imagen de la leyenda de un mapa.
- **WCS (Web Coverage Service):** Es un servicio que permite realizar peticiones de cobertura geográfica a través de la web, normalmente *HTTP*. Las operaciones se pueden codificar como *XML* para *HTTP POST* y *SOAP*, y pares clave-valor (*KVP*) para *HTTP GET* y *RPC (Remote Procedure Call)*. Los datos pueden estar disponibles en varios formatos, como *DTED (Digital Terrain Elevation Data)*, *GeoTIFF*, *HDF-EOS (Hierarchical Data Format – Earth*

Observing System), o *NITFS (National Imagery Transmission Format Standard)*. Permite además añadir rangos de información a las localizaciones.

- **WPS (Web Processing Service):** Proporciona reglas para estandarizar las entradas y salidas de las peticiones y respuestas de servicios de procesamiento geoespacial. Define operaciones para obtener metadatos del servicio, describir entradas y salidas ejecutar el servicio. Se puede usar *HTTP* o *SOAP* codificadas en la *URL (Uniform Resource Locator)* o en *XML*. Normalmente no se invocan directamente, sino desde los controles de la aplicación cliente.

2.3.1.2 JSON, GeoJSON, y TopoJSON

JSON (JavaScript Object Notation) es un formato estándar para datos que destaca por ser ligero y rápido y, por tanto, muy útil para desarrollos web. Los datos en formato *JSON* pueden ser utilizados por prácticamente todos los lenguajes de programación (como *Java*, *C#*, *C*, *C++*, *PHP*, *JavaScript*, *Python*, etc.). A continuación, se muestra un ejemplo de un fichero *JSON* [18]:

```
{
  "marcadores":[
    {
      "latitude": 40.416875,
      "longitude": -3.703308,
      "city": "Madrid",
      "description": "Puerta del Sol"
    },
    {
      "latitude": 40.417438,
      "longitude": -3.693363,
      "city": "Madrid",
      "description": "Paseo del Prado"
    },
    {
      "latitude": 40.407015,
      "longitude": -3.691163,
      "city": "Madrid",
      "description": "Estación de Atocha"
    }
  ]
}
```

Tabla 2-3: Ejemplo de fichero JSON [18]

GeoJSON es un formato estándar *OGC* para codificar colecciones de características geográficas simples junto con sus atributos no espaciales utilizando *JSON*. Los objetos *GeoJSON* pueden representar una geometría, una característica (*feature*) o una colección de características. *GeoJSON* admite los siguientes tipos de geometría: *Point* (ubicaciones puntos de interés, etc.), *LineString* (carreteras, fronteras, etc.), *Polygon* (países, parcelas, etc.), y colecciones de los anteriores (*MultiPoint*, *MultiLineString*, *MultiPolygon* y *GeometryCollection*). Las características (*features*) en *GeoJSON* contienen un objeto geométrico y propiedades adicionales, y se pueden definir también colecciones de características (*FeatureCollection*) [19].

El formato **TopoJSON**, es una extensión de *GeoJSON* que codifica topología geoespacial y proporciona ficheros más compactos que *GeoJSON*.

En la siguiente tabla vemos des ejemplos de fichero *GeoJSON* y *TopoJSON* [20]:

<i>GeoJSON</i>	<i>TopoJSON</i>
<pre>{ type: "Feature Collection" features: [{ "type": "Feature", "geometry": { "type": "Point", "coordinates": [125.6, 10.1] }, "properties": { "name": "Dinagat Islands" } }] }</pre>	<pre>{ "type": "Topology", "arcs": [0: Array(11) 0: Array(2) 0: 17078 1: 14810 1: (2) [72, -46], ...], "transform": { scale: (2) [0.014124293785310738, 0.00818760487616099] translate: (2) [-180, -85.6] }, "objects": { countries: { *Geometry Collection Object* } } }</pre>

Tabla 2-4: Ejemplo de ficheros *GeoJSON* y *TopoJSON* [15]

2.3.1.3 EUMETSAT

Como ya se mencionó al presentar algunos grandes proyectos relacionados con los gemelos digitales oceánicos, *EUMETSAT* (*European Organization for the Exploitation of Meteorological Satellites*) es una fuente de datos de gran importancia. En un apartado posterior se indicará el tipo de información que proporciona, por lo que en este apartado sólo se indicarán los formatos que maneja. En la web de *EUMETSAT* [21], se describe la amplia variedad de formatos y empaquetados de datos que utiliza *EUMETSAT* para la difusión y distribución de productos:

2.3.1.3.1 Formatos de datos

1. **HRIT (Transmisión de Información de Alta Velocidad):** estándar para la difusión de datos de satélites geoestacionarios a usuarios a través de transmisión directa, con velocidades de hasta 10 Mbps. Se definen formatos generales, pero los operadores pueden personalizarlos.
2. **NetCDF (Network Common Data Form):** es un importante formato utilizado en muchos campos científicos como forma estándar de almacenar datos y poder compartirlos de manera eficiente. Las principales ventajas de este estándar son las siguientes:
 - **Abierto:** con interfaces de programación para casi todos los lenguajes.
 - **Autodescriptivo:** incluye información sobre los datos que contiene.
 - **Portable:** soporta diferentes formatos de datos enteros, caracteres y punto flotante.
 - **Escalable:** permite el acceso eficiente a un pequeño subconjunto de datos dentro de un conjunto grande.
 - **Anexable:** permite agregar datos a un archivo *NetCDF* sin copiar el conjunto de datos ni redefinir su estructura.

- **Compatible:** permite a un escritor y varios lectores acceder simultáneamente al mismo archivo *NetCDF*.
 - **Archivable:** en el sentido de compatibilidad de versiones anteriores en versiones actuales y futuras.
3. **HDF (Formato de Datos Jerárquicos):** Diseñado para grandes cantidades de datos organizados jerárquicamente, utilizado en datos ambientales.
 4. **BUFR (Formato Universal Binario para Representación de Datos Meteorológicos):** Estándar binario de la Organización Meteorológica Mundial (*WMO*) para datos de observación. Diseñado para permitir un almacenamiento y un intercambio eficientes.
 5. **GRIB (Formato Binario Estándar de la WMO para Datos en rejilla):** también estándar *WMO* utilizado para la difusión de grandes volúmenes de datos meteorológicos en rejilla.
 6. **XML/CAP (Lenguaje de Marcado Extensible/Protocolo de Alerta Común):** *XML* facilita el intercambio de datos entre sistemas, y *CAP* es un protocolo de alerta mantenido por la Organización para el Avance de los Estándares de Información Estructurada (*OASIS*) y utilizado por *EUMETSAT* para advertencias sobre amenazas a la vida o propiedad.
 7. **McIDAS (Sistema Interactivo de Acceso a Datos por Computadora):** Mantenido por la Universidad de *Wisconsin-Madison*, utiliza un formato de datos patentado para el análisis y visualización de datos meteorológicos, incluidos los satelitales.
 8. **AHRPT (Transmisión Avanzada de Imágenes de Alta Velocidad):** Estándar acordado por *CFMS (The Coordination Group for Meteorological Satellites)* para la difusión de datos de satélites de órbita terrestre baja a velocidades superiores a 500 Kbps.

2.3.1.3.2 Formatos de paquetes

1. **SENTINEL-SAFE:** El formato *SENTINEL-SAFE* está diseñado para gestionar diversos tipos de datos, pudiendo estar en uno o varios archivos. Se clasifica la información en dos categorías principales: Datos y Metadatos. Un producto en este formato contiene un archivo de manifiesto (documento *XML* que define el contenido del paquete), archivos binarios o *XML* (contenidos de datos o metadatos), y archivos de esquema *XML* para representar información binaria. El tamaño de un producto *Sentinel* varía según el tipo de datos que contenga, pudiendo algunos contener varios archivos *netCDF*. Los archivos *netCDF* pueden leerse con bibliotecas en varios lenguajes de programación, y *Sentinel-3 Toolbox* facilita la apertura, visualización, procesamiento y análisis de datos *OLCI (Ocean and Land Color Instrument)* y *SLSTR* (Radiómetro de temperatura de la superficie del mar y de la tierra) de la misión *Sentinel-3*.
2. **Submission Information Package (SIP):** El *SIP* es un archivo *ZIP* comprimido que incluye un archivo de manifiesto (*manifest.xml*) proporcionando información/metadatos sobre el *SIP* y sus elementos, así como metadatos *EOP (Earth Observation Product) (EOPMetadata.xml)* según el perfil de metadatos de observaciones y mediciones de observación de la Tierra de *OGC*. Para atributos específicos de *EUMETSAT*, se amplía de acuerdo con los principios *OGC*. Además, el paquete *SIP* contiene objetos de datos que llevan información del producto, como datos, geolocalización, vista rápida u otros. Los esquemas *XML* correspondientes pueden solicitarse en *EUMETSAT*.

2.3.1.4 Formato SHAPEFILE

Según se puede leer en Wikipedia [22] el formato *ESRI Shapefile (SHP)* es un estándar propietario de datos espaciales desarrollado por *ESRI*, ampliamente utilizado en Sistemas de Información Geográfica. Es un formato vectorial que almacena la ubicación y atributos de elementos geográficos, pero carece de capacidad topológica. Se compone de al menos tres archivos: *.shp* (geometría), *.shx* (índice), y *.dbf* (base de datos de atributos).

Este formato tiene limitaciones, como la gestión de múltiples archivos, un límite de 2 GB en la base de datos, y problemas con consultas espaciales, fechas, y nombres de campos. Aunque sigue siendo popular, se recomienda migrar a alternativas más modernas como *GeoPackage* debido a sus limitaciones y la evolución de otros formatos más avanzados para el almacenamiento de datos geoespaciales.

2.3.1.5 Formato GeoPackage

Tal y como se puede ver en [23], *GeoPackage* es un formato de archivo universal basado en *SQLite* para compartir datos espaciales vectoriales y ráster, siendo una alternativa a formatos como *GeoTIFF* y *shapefile*. La especificación de *GeoPackage* incluye convenciones para almacenar entidades vectoriales, teselas ráster, atributos alfanuméricos y extensiones.

Las principales ventajas de *GeoPackage* son:

1. Compacto, abierto y basado en estándares de la *OGC*, interoperable en diversos entornos.
2. Dispone de índices espaciales para acelerar búsquedas y visualización en *GIS*.
3. Un único archivo “*.gpkg*” facilita la transferencia de información geoespacial.
4. Admite varios tipos de geometrías en un solo archivo.
5. Permite el acceso y actualización directa sin conversiones intermedias.
6. Diseñado para almacenar datos complejos y voluminosos, hasta 140 TB.
7. Soportado por *GDAL* (librería software para la lectura y escritura de formatos de datos geoespaciales) para la conversión de formatos de datos.
8. Compatible con software *GIS* ampliamente utilizado, como *ArcGIS*, *QGIS*, *GeoServer*, entre otros.
9. Útil en dispositivos móviles y entornos con conectividad limitada.

2.3.2 Fuentes de datos institucionales

En este apartado se presentarán las principales fuentes de datos disponibles a nivel institucional tanto de ámbito internacional como nacional y que proporcionan información oceánica y meteorológica ya sea como medidas provenientes directamente de los sensores, como datos procesados o como predicciones obtenidas a partir de modelos. Una vez más, se quiere hacer hincapié en que el objetivo no es entrar en el detalle de los productos, sus formas de acceso o sus formatos concretos, sino dar una idea de lo que existe y donde encontrarlo.

2.3.2.1 Catálogo de fuentes ODIS de IOC

Según se puede leer en [24], “El ‘Catálogo de fuentes’ *ODIS (Oceanic Data and Information System)* de *IOC (Intergovernmental Oceanographic Commission)* pretende ser un catálogo en línea, navegable y con capacidad de búsqueda, de fuentes/sistemas de datos e información existentes relacionados con los océanos, así como de productos y servicios. También proporcionará información sobre productos y visualizará el mapa (entidades y sus conexiones) de datos oceánicos y fuentes de información”. En la *Figura 2-4*, se puede ver la página web del catálogo desde la que se puede acceder a diferentes tipos de información:

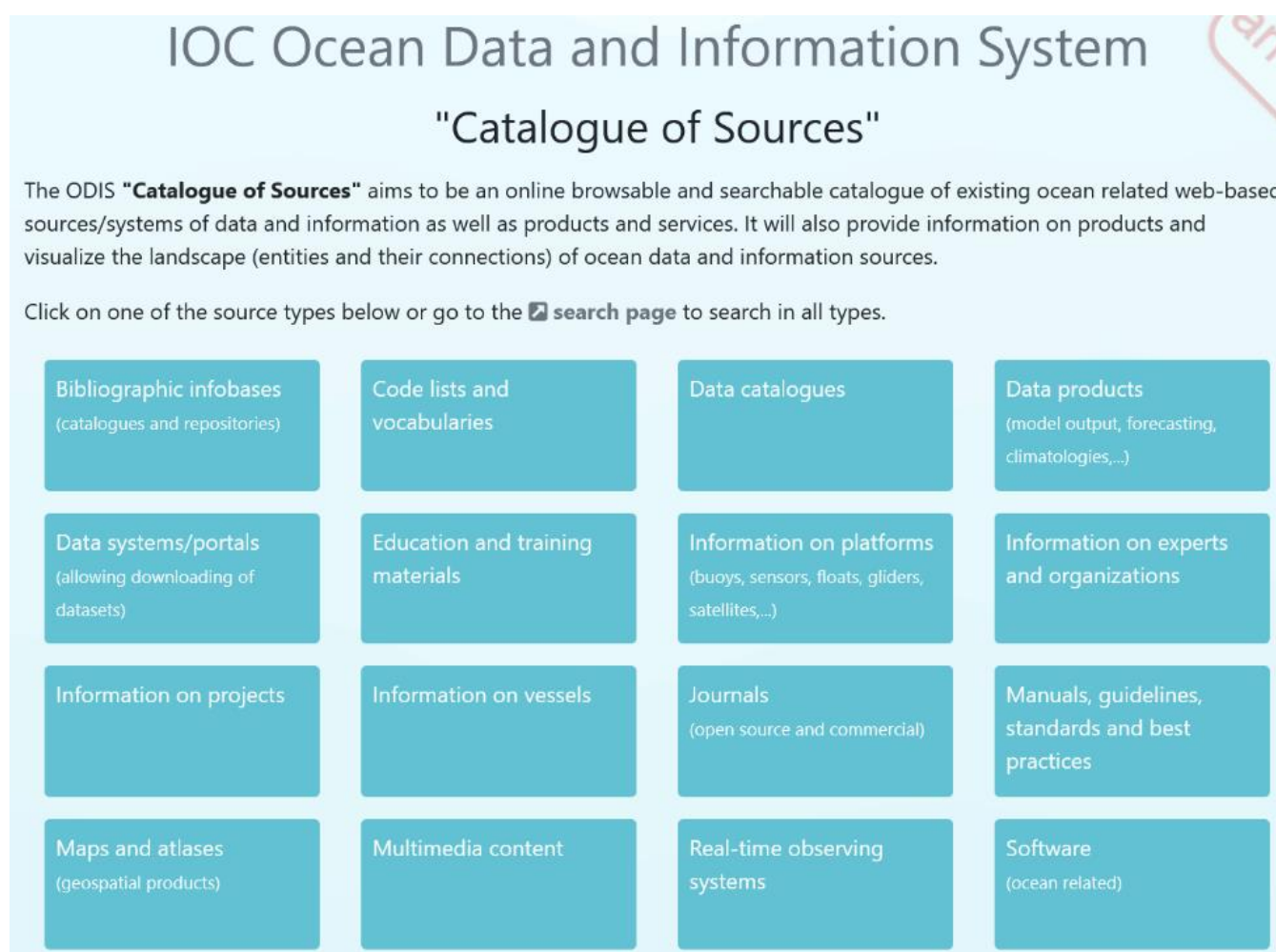


Figura 2-4: Catálogo de fuentes ODIS [24]

Dentro de esta iniciativa, el Programa "Intercambio Internacional de Datos e Información Oceanográficos" (*IODE*) [25], iniciado en 1961, busca mejorar la investigación marina mediante el intercambio de datos entre Estados miembros. Sus objetivos incluyen facilitar el acceso a datos marinos, promover su archivo a largo plazo, y apoyar programas científicos internacionales. La red *IODE* ha recopilado y archivado millones de observaciones oceánicas, ampliando su enfoque a diversas áreas de la oceanografía. Además, el programa se compromete con la accesibilidad y el archivo a largo plazo de datos para proteger los recursos presentes y futuros. El uso de tecnologías de la información y comunicaciones (*TIC*) ha llevado a una transición hacia una red descentralizada de centros de datos, permitiendo una participación más activa de científicos y un acceso más fácil a la información. Este cambio requiere el desarrollo continuo de estándares y aplicaciones.

2.3.2.2 EMODnet

EMODnet (Red Europea de Observación y Datos Marinos), aborda el desafío de la recopilación de datos marinos fragmentados [26]. Su objetivo es proporcionar acceso rápido a información marina confiable para hacer frente a amenazas ambientales, desarrollar políticas protectoras y fomentar el crecimiento económico sostenible. *EMODnet* opera bajo la filosofía de "recopilar una vez y utilizar muchas veces", ahorrando aproximadamente mil millones de euros anuales. La red abarca siete temas, incluyendo batimetría, biología, química, geología, actividades humanas, física y hábitats del lecho

marino. Las pasarelas de *EMODnet* conectan a los usuarios con observaciones estandarizadas y productos de datos procesados, promoviendo la accesibilidad.

EMODnet, está ahora completamente operativo en su cuarta fase, involucrado a más de 120 organizaciones colaboradoras. La iniciativa prioriza los requisitos de los usuarios y colabora con otros proveedores de conocimientos marinos. El Portal de Ingesta de Datos de *EMODnet*, lanzado en 2017, se dirige activamente a conjuntos de datos para llenar vacíos clave. Los principios rectores incluyen recopilar datos una vez, desarrollar estándares, proporcionar financiamiento sostenible y reconocer los datos marinos como un bien público.

Entre los beneficios de *EMODnet* se incluyen facilitar la planificación espacial marina, reducir la incertidumbre en la predicción del comportamiento del mar, mejorar la eficiencia de los operadores offshore, estimular la competencia e innovación en sectores marítimos, así como contribuir a la toma de decisiones efectiva en diversos ámbitos, todo ello aplicable al sector público, privado, comunidad científica y la sociedad en general.

Para el acceso a los datos y metadatos, *EMODnet* proporciona una serie de servicios tanto OGC como no OGC. A continuación, se detallan estos servicios incluyendo un ejemplo de uso, todo ello detallado en su página web [27].

2.3.2.2.1 Servicios de metadatos

El catálogo central de *EMODnet* permite buscar colecciones de metadatos, servicios y objetos de información relacionados con las diferentes temáticas de *EMODnet*. Los catálogos de datos ofrecen un punto final CSW (*Catalog Service for the Web*) a otras aplicaciones cliente para conectarse al servicio y consultar los registros de metadatos disponibles. A continuación, se indican las principales operaciones disponibles con este servicio:

- **GetCapabilities:** permite a los clientes CSW recuperar metadatos de servicio de un servidor. La respuesta será un documento XML con metadatos de servicio sobre el servidor.

Thematic	CSW GetCapabilities
Central Catalogue	https://emodnet.ec.europa.eu/geonetwork/emodnet/eng/csw?service=CSW&request=GetCapabilities&VERSION=2.0.2

Tabla 2-5: *EMODNet*: Solicitud de metadatos de servicio [27]

- **GetRecords:** permite acceder y filtrar los registros como en el siguiente ejemplo:

```
https://emodnet.ec.europa.eu/geonetwork/emodnet/eng/csw?REQUEST=GetRecords&SERVICE=CSW&VERSION=2.0.2&ELEMENTSETNAME=summary&OUTPUTSCHEMA=http://www.opengis.net/cat/csw/2.0.2&CONSTRAINTLANGUAGE=FILTER&CONSTRAINTLANGUAGE_VERSION=1.1.0&RESULTTYPE=results&TYPENAMES=csw:Record&CONSTRAINT=<ogc:Filter>

xmlns:ogc="http://www.opengis.net/ogc"><ogc:PropertyIsEqualTo><ogc:PropertyName>dc:type</ogc:PropertyName><ogc:Literal>dataset</ogc:Literal></ogc:PropertyIsEqualTo></ogc:Filter>&maxRecords=10
```

Tabla 2-6: *EMODNet*: Solicitud de registros [27]

- **GetRecordById:** se puede usar para obtener los registros del catálogo de metadatos usando su identificador. Ejemplo:

<https://emodnet.ec.europa.eu/geonetwork/emodnet/eng/csw?request=GetRecordById&service=CSW&version=2.0.2&elementSetName=full&id=5c0f13ee-098a-442c-abfc-d4c08a527476>

Tabla 2-7: *EMODNet*: Solicitud de registros mediante identificador [27]

2.3.2.2.2 Servicios de datos

Las capas de datos de *EMODnet* están disponibles como *Web Feature Service (WFS)* o *Web Coverage Service (WCS)* según los estándares *OGC*.

Para ciertas temáticas también se proporcionan servicios web no *OGC* como, por ejemplo, un servidor central *ERDDAP (Environmental Research Division's Data Access Program)*, un servicio *REST* para *EMODnet Bathymetry DTM (Digital Terrain Model)*, uso de parámetros específicos en solicitudes *WFS* para *EMODnet Biology*, y un servicio *OPeNDAP (Open-source Project for a Network Data Access Protocol)* para *EMODnet Chemistry*.

A continuación, se describe cada servicio proporcionando el punto de acceso según corresponda:

- **Web Feature Service (WFS):** estándar para intercambiar datos vectoriales consultando tanto la estructura de datos como los datos de origen. Las operaciones básicas son *GetCapabilities*, *DescribeFeatureType* y *GetFeature*. Admite una variedad de formatos de salida *WFS* (Ej.: *GML*, *shapefile*, *JSON*, *GeoJSON*, *CSV*, ...).

En la siguiente tabla se muestran los datos *WFS* disponibles y los puntos de acceso:

Thematic	Description	WFS GetCapabilities
Bathymetry	Data Products	https://ows.emodnet-bathymetry.eu/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Biology	Data Products	https://geo.vliz.be/geoserver/Emodnetbio/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Biology	Occurrence data	https://geo.vliz.be/geoserver/Dataportal/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Chemistry	Litter	https://sextant.ifremer.fr/services/wfs/emodnet_chemistry2?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Chemistry	Contaminants	https://nodc.ogs.it/geoserver/Contaminants/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Chemistry	CDI data discovery and access service	https://geo-service.maris.nl/emodnet_chemistry/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Chemistry	Distribution of CDI observations per data category (P36) and MSFD sea regions	https://geo-service.maris.nl/emodnet_chemistry_p36/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
Geology	Sea-floor (bedrock)	https://drive.emodnet-geology.eu/geoserver/bgr/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0

<i>Thematic</i>	<i>Description</i>	<i>WFS GetCapabilities</i>
<i>Geology</i>	<i>Marine Minerals</i>	https://drive.emodnet-geology.eu/geoserver/gsi/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Geology</i>	<i>Seabed substrate maps</i>	https://drive.emodnet-geology.eu/geoserver/gtk/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Geology</i>	<i>Events and Probabilities</i>	https://drive.emodnet-geology.eu/geoserver/ispra/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Geology</i>	<i>Coastal Behaviour</i>	https://drive.emodnet-geology.eu/geoserver/tno/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Geology</i>	<i>Submerged Landscapes</i>	https://drive.emodnet-geology.eu/geoserver/bgs/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Geology</i>	<i>Index of borehole and geophysics data</i>	https://drive.emodnet-geology.eu/geoserver/geus/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Human Activities</i>	<i>Data and Data Products</i>	https://ows.emodnet-humanactivities.eu/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Physics</i>	<i>Data and Data Products</i>	https://prod-geoserver.emodnet-physics.eu/geoserver/ows?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Seabed Habitats</i>	<i>General datasets and products</i>	https://ows.emodnet-seabedhabitats.eu/geoserver/emodnet_open/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0
<i>Seabed Habitats</i>	<i>Individual habitat map and model datasets</i>	https://ows.emodnet-seabedhabitats.eu/geoserver/emodnet_open_maplibrary/wfs?SERVICE=WFS&REQUEST=GetCapabilities&VERSION=2.0.0

Tabla 2-8: EMODNet: Servicios WFS [27]

Dentro de este servicio se quiere destacar la temática de Actividades Humanas (*Human Activities*), ya que en esta capa se publica información relacionada con las infraestructuras críticas, tales como instalaciones de producción de energía, cables de comunicaciones, áreas protegidas, polígonos o áreas militares, zonas de vertidos de aguas, etc., que podrían ser tenidas en cuenta para planificar las misiones o para anticipar posibles zonas de riesgo, amenazas o catástrofes.

- **Servicio de cobertura web (WCS):** permiten solicitudes de datos de cobertura (ráster) o productos de datos en rejilla. La siguiente tabla muestra la información disponible y su enlace:

<i>Thematic</i>	<i>Description</i>	<i>WCS GetCapabilities</i>
<i>Bathymetry</i>	<i>Data Products</i>	https://ows.emodnet-bathymetry.eu/wcs?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1
<i>Biology</i>	<i>Data Products</i>	https://geo.vliz.be/geoserver/Emodnetbio/wcs?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1
<i>Biology</i>	<i>New Data Products</i>	https://ows.emodnet.eu/geoserver/biology/ows?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1
<i>Human Activities</i>	<i>Data and Data Products</i>	https://ows.emodnet-humanactivities.eu/wcs?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1
<i>Seabed Habitats</i>	<i>Individual habitat map and model datasets</i>	https://ows.emodnet-seabedhabitats.eu/geoserver/emodnet_open_maplibrary/wcs?SERVICE=WCS&REQUEST=GetCapabilities&VERSION=2.0.1

Tabla 2-9: EMODNet: Servicios WCS [27]

- **Servicios No OGC**

Los servicios web No-OGC proporcionados son los mostrados en la tabla siguiente:

<i>Thematic</i>	<i>Web Service</i>	<i>URL</i>
<i>EMDOnet</i>	<i>ERDDAP</i>	https://erddap.emodnet.eu/erddap/index.html
<i>Bathymetry</i>	<i>REST API</i>	https://rest.emodnet-bathymetry.eu/
<i>Chemistry</i>	<i>THREDDS</i>	http://opendap.oceanbrowser.net/thredds/catalog/data/emodnet-domains/catalog.html XML version: http://opendap.oceanbrowser.net/thredds/catalog/data/emodnet-domains/catalog.xml
<i>Physics</i>	<i>ERDDAP</i>	https://prod-erddap.emodnet-physics.eu/erddap/index.html

Tabla 2-10: EMODNet: Servicios No OGC [27]

2.3.2.3 Copernicus

El ecosistema de espacio de datos de *Copernicus* marca un avance significativo en la evolución de los datos de observación de la Tierra, proporcionando acceso inmediato y gratuito a grandes cantidades de información de los satélites *Copernicus Sentinel* y misiones contribuyentes. Diseñado para mejorar la accesibilidad y explotación de estos datos, el servicio busca facilitar la creación de aplicaciones cruciales para un futuro sostenible. Con escalabilidad, facilidad de uso y potentes capacidades de procesamiento, se proyecta como la plataforma de referencia para extraer conocimientos de los datos de *Copernicus*.

El servicio ofrece un fácil descubrimiento, visualización, descarga y análisis de datos de observación de la Tierra, con acceso a herramientas de procesamiento para diversas aplicaciones. Además, se presenta como una propuesta abierta para nuevos servicios futuros. Con el objetivo principal de garantizar la disponibilidad instantánea de datos, el servicio incluirá el archivo completo de datos de los satélites *Copernicus Sentinel*, así como misiones contribuyentes, accesibles de forma gratuita [28].

El ecosistema se destaca por ser el primer entorno que proporciona acceso a casi todos los datos históricos de las constelaciones de satélites *Sentinel*, llenando un vacío importante. Los usuarios pueden beneficiarse de una amplia gama de herramientas y servicios para el procesamiento, análisis y extracción de conocimientos de datos. Se abordan específicamente las misiones *Sentinel-1*, *Sentinel-2*, *Sentinel-3* y *Sentinel-5P*, cada una con objetivos y capacidades únicas para monitorear cambios en la Tierra, condiciones de la superficie terrestre, topografía del mar, temperatura y composición atmosférica.

En la dirección web específica de datos marítimos [3] se pueden ver diferentes tipos de datos que se pueden descargar. Entre ellos cabe destacar para el interés de este TFM los siguientes: altura del nivel del mar, densidad, salinidad, temperatura, velocidad del agua, viento, olas y capa de hielo.

Una forma de ver todo lo que hay disponible para descarga es entrar, previo registro, en la página del catálogo desde la web mencionada, que tiene la apariencia que se muestra en la *Figura 2-5*:

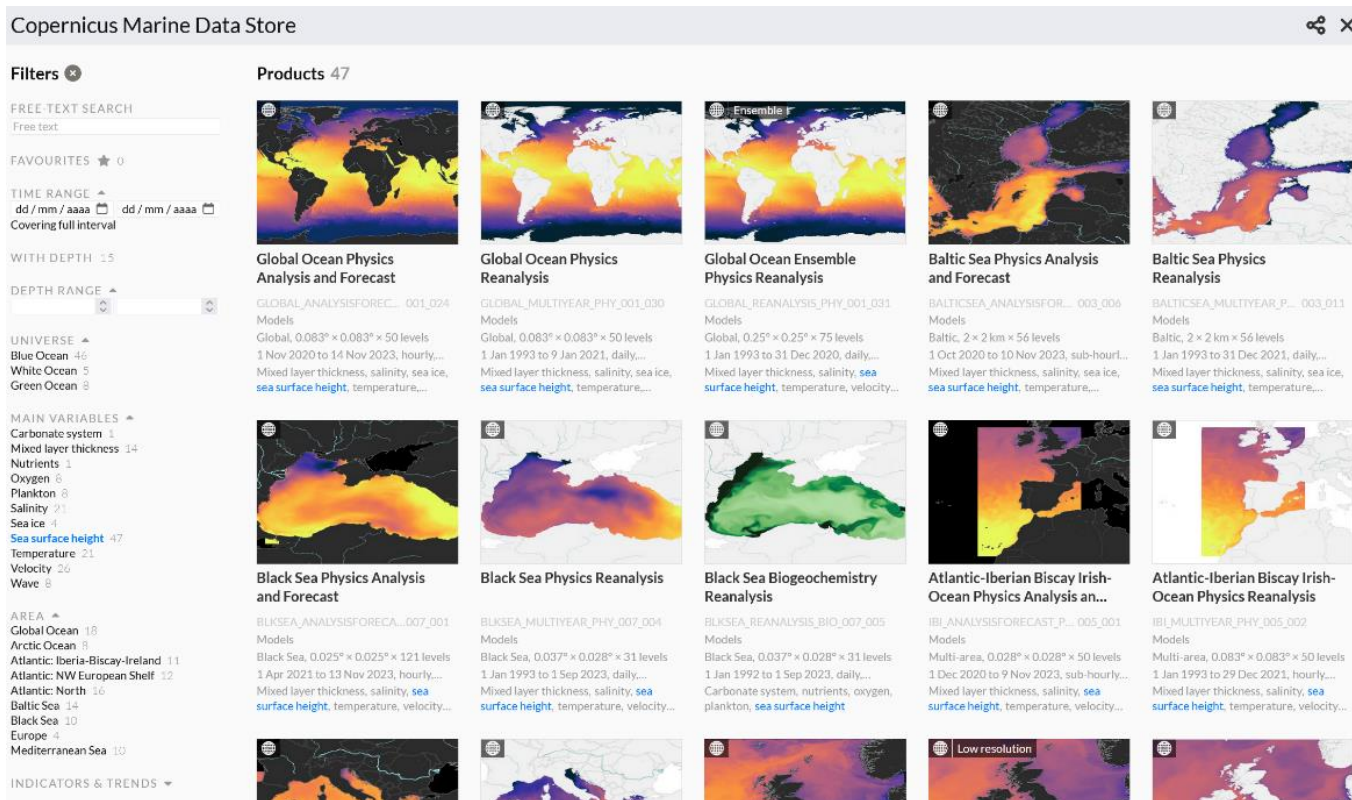


Figura 2-5: Catálogo de datos de Copernicus [3]

Si, por ejemplo, se quieren descargar los datos de temperatura de la superficie del mar, en la propia ventana de añadir la capa se pueden obtener los comandos para automatizar la descarga de esos datos (botón “Automate”), tal y como se muestra en la *Figura 2-6*:

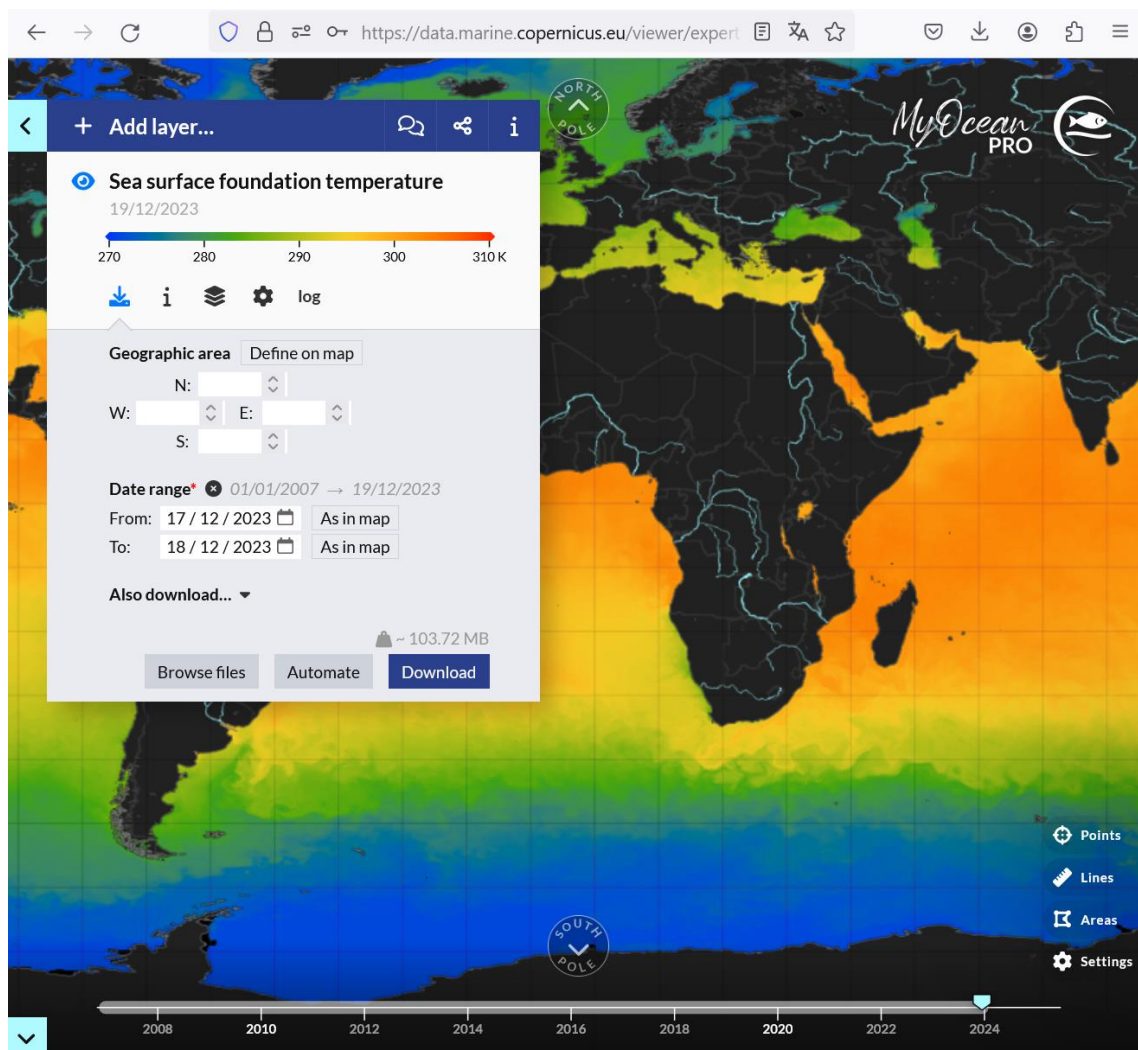


Figura 2-6: Ejemplo de comando de descarga de datos a través de API [3]

Por ejemplo, para descargar los datos mencionados se podría hacer a través de dos interfaces de automatización, mediante interfaz de línea de comandos (CLI) y desde Python:

Command Line Interface (CLI)

To download this subset programmatically in CLI, run the following command:

```
copernicus-marine subset --dataset-id METOFFICE-GLO-SST-L4-NRT-OBS-SST-V2 --variable analysed_sst  
--start-datetime 2023-12-17T00:00:00 --end-datetime 2023-12-18T23:59:59 --minimum-longitude -180 --  
maximum-longitude 180 --minimum-latitude -90 --maximum-latitude 90
```

Python Application Programming Interface (API)

To download this subset from a Python environment, run this snippet:

```
import copernicus_marine_client as copernicusmarine
```

```

copernicusmarine.subset(
  dataset_id="METOFFICE-GLO-SST-L4-NRT-OBS-SST-V2",
  variables=["analysed_sst"],
  minimum_longitude=-180,
  maximum_longitude=180,
  minimum_latitude=-90,
  maximum_latitude=90,
  start_datetime="2023-12-17T00:00:00",
  end_datetime="2023-12-18T23:59:59",
)

```

Tabla 2-11: Acceso automatizado a datos de Compernicus [3]

2.3.2.4 Agencia Española de Meteorología (AEMET)

AEMET OpenData [29] proporciona un *API REST* a través de la cual se pueden descargar gratuitamente los datos indicados en el Anexo II de la Resolución de 30 de diciembre de 2015 de *AEMET* [30], por la que se establecen los precios públicos que han de regir la prestación de servicios meteorológicos y climatológicos, publicada en el BOE n.º 4 de 5 de enero de 2016. Para poder realizar peticiones, es necesario introducir una clave de *API* (*API Key*).

A modo de resumen, *AEMET* proporciona servicios de observación y de predicción.

- Entre los servicios de observación proporciona información en forma de datos, mapas o imágenes procedentes de diferentes redes o fuentes:
 - Observación convencional: temperatura, viento, precipitaciones, etc.
 - Redes especiales: radiación, ozono, contaminación.
 - Red de detección de rayos.
 - Rede de Satélites.
 - Red de radares.
- Como servicios de predicción, ofrece modelos numéricos de predicción para diferentes variables meteorológicas como precipitaciones, viento, temperatura, etc. Puede proporcionar datos de cada variable, indicando número de pasadas del modelo, resolución, alcance temporal, etc. Estas predicciones también se ofrecen como productos gráficos.

En el Anexo III de la mencionada resolución se indican, así mismo, las exclusiones al pago de los precios públicos, donde aparece explícitamente la exclusión para en el caso de apoyo meteorológico prestado a las Fuerzas Armadas, la Defensa Nacional y a las Fuerzas y Cuerpos de Seguridad del Estado, Autonómico y Local para el ejercicio de sus actividades, así como la navegación aérea, avisos y predicciones meteorológicas para las autoridades de protección civil en los diferentes ámbitos nacional, autonómico y local, y también en salvamento marítimo y situaciones de emergencia.

2.3.2.5 Puertos del estado

Puertos del Estado proporciona en su sitio web un servicio online de descarga gratuita de datos océano-meteorológicos, mediante el cual contribuye a un mejor conocimiento del litoral nacional, así como al desarrollo y gestión de las actividades costeras y portuarias.

A través de este servicio se pueden descargar series temporales de datos oceanográficos y meteorológicos procedentes de los sistemas de monitorización del medio marino de Puertos del Estado. Se puede acceder a este servicio de descargas a través de su página web [31].

La Figura 2-7 muestra la interfaz web ofrecida por la institución, donde la descarga se haría desde el menú gráfico titulado Datos Históricos.

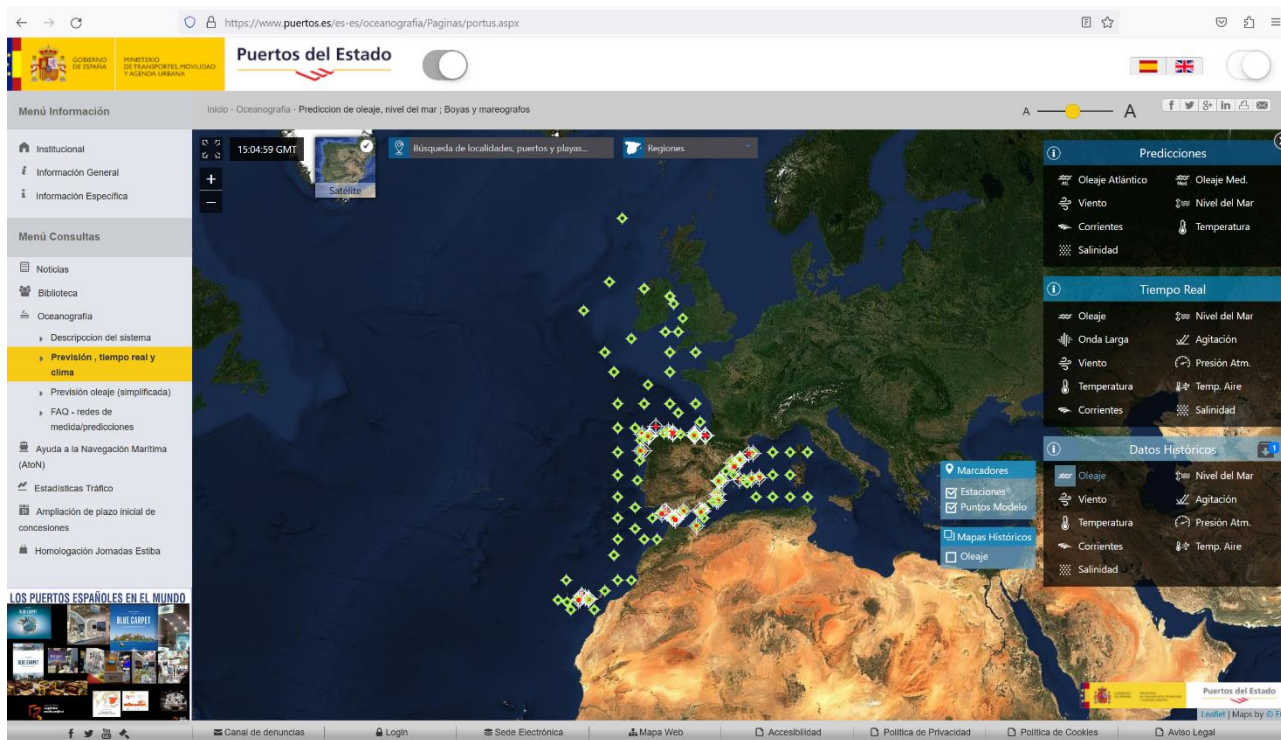


Figura 2-7: Página de Puertos del Estado para visualización y descarga de datos [31]

Según se indica en el manual de usuario [32], el formato de los ficheros de datos es .csv (*comma separated values*) con tabuladores como separadores de columnas y ordenados por la primera columna que corresponde a la fecha.

Actualmente, no hay disponible un *API* para el acceso a las descargas de los datos, pero tras contactar con la institución, indican que se puede solicitar una conexión *SFTP* (*Secure File Transfer Protocol*) para acceder a los datos obtenidos a partir de su red de medidores.

2.3.2.6 MeteoGalicia

MeteoGalicia es un organismo meteorológico dependiente de la *Xunta* de Galicia que presta servicios de observación y predicción.

MeteoGalicia genera diariamente grandes volúmenes de datos meteorológicos en ficheros de cientos de *Megabytes* o *Gigabytes*. El servidor *Thredds* (*Thematic Realtime Environmental Distributed Data Service*) ofrece diversos servicios de descarga, como *HTTPServer*, *OpenDAP*, *NetCDFSubset*, *WMS* y *WCS*, permitiendo a usuarios acceder a predicciones atmosféricas y oceánicas [33 y 34]. Estos servicios facilitan la descarga directa o la extracción de subsecciones específicas, optimizando el uso de recursos, y permitiendo la integración de datos en diferentes aplicaciones de análisis de datos.

En la referencia [34] se dispone de un catálogo de datos y predicciones de los modelos disponibles en dicho servidor *THREDDS*:

2.3.2.6.1 Datos operacionales del modelo

- **WRF (Weather Research Forecast) 1Km**
 - Variables del modelo a nivel de superficie: Archivos *NetCDF* con variables de superficie del modelo *WRF*. El modelo *WRF* se ejecuta operativamente una vez al día, se inicializa a las 00 UTC (*Universal Time Coordinated*) y dura 96 horas. Se configuran dos dominios anidados para una resolución de 5 km y 1 km.
- **WRF (Weather Research Forecast)**
 - **Variables del modelo a nivel de superficie:** archivos *NetCDF* con variables de superficie del modelo *WRF*. El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00 UTC y 12 UTC, el primero funciona durante 96 horas y el segundo durante 84 horas. Se configuran tres dominios anidados para una resolución de 36 km, 12 km y 4 km.
 - **Modelo de salida directa/sin procesar:** archivos de salida del modelo *NetCDF WRF* con variables de pronóstico del modelo. El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00 UTC y 12 UTC, el primero funciona durante 96 horas y el segundo durante 84 horas. Se configuran tres dominios anidados para una resolución de 36 km, 12 km y 4 km. Estos archivos se utilizarán generalmente para anidar otros modelos.
 - **Variables de modelo de diagnóstico 3D y de superficie:** Archivos *Grib* con superficie pronosticada y variables verticales. El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00 UTC y 12 UTC, el primero funciona durante 96 horas y el segundo durante 84 horas. Tres dominios anidados están configurados para una resolución de 36 km, 12 km y 4 km.
- **WRF 1.3Km**
 - **Variables de modelo de diagnóstico 3D y de superficie:** Archivos *Grib* con variables de superficie y verticales. El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00UTC y 12UTC, el primero funciona durante 72 horas y el segundo durante 84 horas. Hay tres dominios disponibles: *d04* para la región de *Artabro*, *d05* para la región de *Rias Baixas* y *d06* para la región Norte de Portugal.
 - **Archivos de salida del modelo NetCDF WRF con variables de pronóstico del modelo.** El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00UTC y 12UTC, el primero funciona durante 72 horas y el segundo durante 84 horas. Hay tres dominios disponibles: *d04* para la región de *Artabro*, *d05* para la región de *Rias Baixas* y *d06* para la región del Norte de Portugal. Modelo Salida bruta/directa.
 - **Variables de modelo de diagnóstico 3D y de superficie:** Archivos *NetCDF* con variables de superficie del modelo *WRF*. El modelo *WRF* se ejecuta operativamente dos veces al día inicializado a las 00UTC y 12UTC, el primero funciona durante 72 horas y el segundo durante 84 horas. Hay tres dominios disponibles: *d04* para la región de *Artabro*, *d05* para la región de *Rias Baixas* y *d06* para la región Norte de Portugal.
- **WW3 (Wave Watch III)**
 - **Parámetros de oleaje calculados a partir de espectros de potencia pronosticados:** archivos *NetCDF* con los parámetros de oleaje más relevantes. Estas variables se calculan a partir de los espectros de potencia pronosticados por el modelo *WW3*. El modelo *WW3* funciona operativamente dos veces al día inicializado a las 00 UTC y 12 UTC, el primero

funciona durante 96 horas y el segundo durante 84 horas. Se ejecutan tres dominios anidados para diferentes resoluciones.

- **SWAN (*Simulating Waves Nearshore*)**
 - **Parámetros de olas calculados a partir de espectros de potencia pronosticados:** archivos *NetCDF* con parámetros de olas. Estas variables se calculan a partir de los espectros de potencia pronosticados por el modelo SWAN. El modelo SWAN se ejecuta operativamente una vez al día inicializado a las 12UTC y está obligado por el modelo operativo WW3 de *Meteogalicia*. El tiempo de entrega de pronóstico es de 84 horas para proporcionar pronóstico de 3 días y se utilizan dos dominios para pronosticar las regiones geográficas de Artabro y Rias Baixas.
- **ROMS (*Regional Ocean Modeling System*)**
 - **Salida del modelo interpolada a capas de profundidad fija:** archivos *NetCDF* con variables oceánicas a profundidades fijas. Estas variables se derivan del resultado crudo del modelo ROMS. Este modelo se ejecuta operativamente una vez al día inicializado a las 00UTC y es forzado por las salidas de los modelos WRF y SWAT de *MERCATOR* y *Meteogalicia*. Tiene una duración de 96 horas para proporcionar previsiones de 4 días en un único dominio de resolución de 2 km que cubre el noroeste de la Península Ibérica.
 - **Salida sin procesar del modelo:** archivos *NETCDF*, salida original del modelo ROMS. ROMS se ejecuta operativamente una vez al día a las 00UTC, está forzado por *MERCATOR* y el modelo WRF y SWAT de *Meteogalicia*. Funciona para proporcionar previsiones de 4 días en un único dominio de resolución de 2 km que cubre el noroeste de la Península Ibérica.
- **MOHID (Modelo Hidrodinámico)**
 - **Variables oceánicas a diferentes profundidades del modelo:** archivos *NetCDF* con las variables oceánicas más relevantes interpoladas a profundidades fijas. MOHID funciona operativamente una vez al día y está forzado por los modelos ROMS, WRF y SWAT de *MetoGalicia*. Funciona para proporcionar un pronóstico de 2 días para varios dominios de alta resolución.
 - **Salida sin procesar del modelo:** archivos *HDF5*, salida original del modelo MOHID. MOHID funciona operativamente una vez al día a las 09:30UTC, está forzado por los modelos ROMS, WRF y SWAT de *MetoGalicia*. Funciona para proporcionar un pronóstico de 2 días para varios dominios de alta resolución.
- **SWAT (Herramienta para la Evaluación del Suelo y Agua)**
 - **Predicción de caudales y temperaturas de los ríos:** archivos *NetCDF* con caudales de los ríos previstos calculados con la herramienta de evaluación del agua del suelo (SWAT). El modelo SWAT se ejecuta operativamente una vez al día y está obligado por el modelo operativo WRF de *Meteogalicia*. Proporcionan caudal y temperatura media diaria de diferentes ríos gallegos. Estos caudales están calculados para forzar los dominios de alta resolución del MOHID y es por eso que la mayoría de los ríos pertenecen al Espacio Atlántico.
- **CHIMERE**
 - **Salida del modelo del modelo de pronóstico químico meteorológico CHIMERE en formato NetCDF:** cada archivo contiene pronósticos de 72 horas inicializados a las 03

UTC del día indicado. El modelo se ejecuta en dos dominios: un dominio grueso que cubre el norte de la Península Ibérica y otro más fino centrado en Galicia. Se proporcionan resultados del modelo para una amplia gama de especies químicas, así como para el índice de calidad del aire utilizado actualmente por el gobierno gallego. En los límites laterales del dominio grueso, *CHIMERE* se alimenta de pronósticos *C-IFS* de *ECMWF/Copernicus*.

2.3.2.6.2 Datos históricos

- **RAIA batimetría**
 - **Batimetría de alta resolución para el Noroeste de la Península Ibérica:** Archivo *NetCDF* con la Batimetría para el Noroeste de la Península Ibérica. Se utilizaron diferentes fuentes de datos: *GEBCO* (*The General Bathymetric Chart of the Oceans*) como campo de fondo y diferentes mediciones y batimetrías. Todos los datos se fusionaron mediante interpolación óptima.
- **Archivos históricos WRF**
 - **Variables del modelo a nivel de superficie y algunas presiones y variables en alturas específicas cerca de la superficie:** Archivos *NetCDF* para el modelo *WRF* con variables de superficie y verticales en niveles de presión y alturas cerca de la superficie. Los archivos correspondientes a la salida 00UTC para los tres dominios anidados se ejecutan con una resolución de 36 km, 12 km y 4 km.

2.3.2.6.3 Radar

- **Radar**
 - *NetCDF* con valores de reflectividad tomados por el producto *PPI* a 0,5° de elevación.: Los escaneos se realizan a intervalos de 5 minutos. El radar está ubicado en 42,61° N, 8,19° W 750 m a. s. l.

2.3.2.6.4 Datos externos

- **Modelo GFS (Global Forecast System, NCEP)**
 - **Archivo de datos individuales, que componen la colección de ejecución de modelos de pronóstico:** archivos *Grib* con los pronósticos del modelo *GFS*. Sólo algunos de los niveles de presión relevantes están disponibles.
- **Termosalinómetro (Instituto Español de Oceanografía)**
 - **Datos semanales del buque Navaz - IEO:** Datos semanales medidos en la costa gallega del buque *Navaz* del *IEO* (Instituto Español de Oceanografía).

2.3.3 Fuentes de datos propias de la plataforma desplegada

En la arquitectura del Gemelo Digital del Entorno Operativo Marítimo, además de los datos procedentes de fuentes externas institucionales, que en principio solo serán accesibles por la infraestructura central no desplegada en la que se llevará a cabo el entrenamiento del gemelo digital, se recibirán también los datos locales procedentes de las plataformas desplegadas. Estas recolectarán información a partir de la instrumentación y sensores instalados en la propia plataforma y en las

plataformas “auxiliares” desplegables desde esta, tales como vehículos aéreos o submarinos no tripulados (*UAV* o *USV* respectivamente), etc.

Existen soluciones de instrumentación meteorológica y oceánica instalable en las plataformas marinas capaces de proporcionar información muy interesante para alimentar al gemelo digital del entorno, como las que se indican a continuación extraídas de páginas web de fabricantes y que se acompañan de imágenes ([35]-[36]):

- Medidores de la velocidad y dirección del viento, a través de tecnologías sónicas, ultrasónicas, láser, cazoletas, etc.). Existen sensores con conectividad inalámbrica *4G* capaces de proporcionar medidas a distancias de 360° y 10 km de distancia como el de *Vaisala*.



Figura 2-8: Medidores de velocidad y dirección del viento. A la Izquierda uno de la empresa Campbell Scientific y a la derecha uno más sofisticado de la empresa Vaisala [35] y [36]

- Medidores de presión atmosférica, temperatura y humedad



Figura 2-9: Sensor de temperatura y humedad relativa HygroVUE10 de Campbell Scientific [36]

- Medidores de visibilidad mediante sensores basados en la dispersión directa *IR*.



Figura 2-10: Medidor de visibilidad de Campbell Scientific [36]

- Medidores de tiempo actual (meteorológico), capaces de medir mediante tecnología láser el tamaño y velocidad de las gotas, determinar la visibilidad, la intensidad y el tipo de precipitación: llovizna, lluvia, nieve, granizo, niebla, rocío, etc.



Figura 2-11: Medidor de tiempo actual [35]

- Medidores de altura y cobertura de nubes mediante tecnología *LIDAR* y detección óptica de nubosidad baja como el de la siguiente figura:

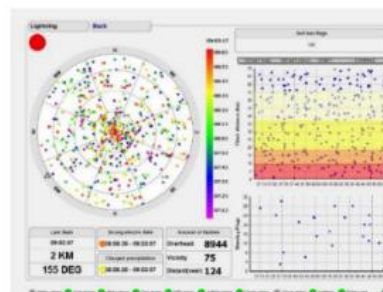


Figura 2-12: Vaisala Ceilometer CL31 [35]

- Detectores de relámpagos como el que se puede ver en la siguiente figura.



BT-D-350



Software module OMC-DOL-BTD with BT-D-350



OMC-140 Multifunctional TFT Display with BT-D-350

Figura 2-13: Detector de relámpagos BT-D-350 de Observator Instruments [37]

Además de esta instrumentación de tipo ambiental, las plataformas navales llevarán instalados un gran número de diferentes sensores para la navegación tales como receptor *GPS*, girocompás, sensores de inclinación y balanceo, cámaras ópticas o *IR* (infrarrojo) que capturan imágenes del entorno que, adecuadamente procesadas, pueden proporcionar diferentes tipos e información ambiental, etc.

Respecto a los protocolos de comunicaciones usados en las plataformas marinas, la mayoría de los sensores instalados en plataformas marinas serán normalmente compatibles con los estándares de la Asociación Nacional de Electrónica Marina (*NMEA*) de comunicación para productos electrónicos marinos, destacando principalmente dos: *NMEA0183* y su evolución *NMEA2000* [34].

NMEA0183 es una combinación de especificación eléctrica y de datos para la comunicación entre los equipos electrónicos marítimos. Opera a diferentes velocidades (4800, 38400, 115200 baudios).

NMEA2000 es un estándar *plug-and-play* creado para facilitar la comunicación entre productos electrónicos marinos, aumentando la velocidad y simplificando la instalación. A pesar del crecimiento de *NMEA2000*, la mayoría de los nuevos dispositivos aún mantienen conectividad *NMEA0183*.

Más recientemente se ha introducido *Signal K* [38], una plataforma de código abierto que supone una revolución en las interfaces electrónicas marinas. Diseñada para intercambiar datos no solo entre instrumentos de una embarcación, sino también entre múltiples embarcaciones, puertos y marinas. *Signal K* se adapta a aplicaciones web y móviles, conectando embarcaciones al "Internet de las cosas". Aunque *NMEA* persiste, *Signal K* ofrece una solución avanzada para las demandas actuales. No obstante, en el ámbito de la Defensa puede no ser una opción muy adecuada por cuestiones de seguridad.

Además de los sensores anteriores, que se podrían instalar en cualquier plataforma marítima de tamaño medio, en plataformas de mayor envergadura o destinadas a misiones de carácter más operativo que comercial o recreacional y, sobre todo, en el ámbito de Defensa, se encontrarán unos sistemas de sensores mucho más sofisticados (y mucho más costosos) como son los sonares y radares de última generación. Estos sistemas, aunque tienen un eminente propósito táctico centrado en la detección de objetos, en función de sus características, podrían proporcionar también diferentes tipos de información

oceanográfica tales como oleaje y corrientes de superficie a través del radar, o corrientes a diferentes profundidades mediante sistemas de tipo sonar con capacidad ACSP (*Acoustic Doppler Current Profiler*) o DVL (*Doppler Velocity Log*) como por ejemplo el indicador de corrientes *Furuno CI-68* [39], o el sistema de mapeo y detección de oleaje *WaveVision 2* de la empresa *Rutter* [40].

También en estas plataformas más grandes y de propósito más operativo, es donde se contempla la posibilidad de que, desde ellas, se despliegan plataformas auxiliares con propósitos específicos de mejora de la conciencia situacional o de naturaleza más operativa, que a su vez pueden proporcionar información útil tanto para la operación como para alimentar al gemelo digital. Este tipo de plataformas auxiliares pueden ser vehículos tripulados o de operación remota, ya sean aéreos, submarinos o de superficie, boyas o sensores remolcados, etc. Normalmente, dispondrán de sensores más simples y reducidos que proporcionarán valores para algunas de las variables indicadas (presión atmosférica, temperatura, etc.), pero con la importante ventaja de poder alejarse de la plataforma principal y recoger datos que no están al alcance de esta última, lo que permitiría ampliar el área de modelización, o porque sean plataformas auxiliares submarinas o voladoras que pueden acceder a diferentes alturas y profundidades donde le es imposible llegar a la plataforma principal.

2.4 Gestión y almacenamiento de los datos

Llegados a este punto, no hace falta decir que la tecnología de gemelo digital requiere de la existencia de unos activos indispensables sin los cuales esta tecnología no tiene ningún sentido: los datos. El gemelo digital se alimenta de datos relevantes sobre el gemelo real que se reciben de forma continua y que permiten controlar la evolución del modelo y proporcionar predicciones y decisiones de calidad basadas en evidencias. Cuantos más datos haya disponibles y que proporcionen información relevante sobre gemelo real, más preciso podrá ser el gemelo digital y con él los productos que se obtengan en su operación: estado exacto del gemelo físico, respuesta en casos hipotéticos, predicciones, etc.

En secciones anteriores ya se ha abordado la cuestión de donde se pueden obtener estos datos para el alcance del trabajo: fuentes externas institucionales y sensores de las plataformas, y también se han tratado los formatos en los que se pueden encontrar los datos, que como se vio eran bastante heterogéneos. La cuestión ahora es la siguiente: ¿cómo se almacenan, gestionan, y explotan esos datos? Para responder a esta cuestión, primero se debe responder a la pregunta de cuál es el volumen de datos que se van a manejar. La respuesta en el contexto de gemelos digitales es fácil: una cantidad ingente de datos y de naturaleza muy diversa. Las fuentes institucionales publican datos constantemente, tanto datos en crudo como datos procesados y datos de predicciones, en formatos de tipo registro o imágenes. Igualmente, las plataformas van a generar también un volumen enorme de datos provenientes de todos los sensores que constantemente proporcionan valores de las variables medidas, como valores registrados, o imágenes, sonidos, etc. Para poder explotar toda esta información y convertirla en valor, hará falta, por un lado, almacenarla con el propósito de entrenar posteriormente los modelos que constituyan el gemelo digital, pero, por otro lado, el gemelo digital también deberá ser capaz de recibir y actualizar su estado con la ingente información que llega en tiempo real, lo que complica aún más la problemática.

Esto que se ha descrito es un claro escenario de lo que se denomina *Big Data*, un concepto omnipresente en un mundo actual donde las organizaciones se han dado cuenta del valor de los datos, y para cuya problemática se han desarrollado sofisticadas soluciones que a su vez están evolucionando de forma constante.

Teniendo en cuenta lo anterior y el objetivo de este trabajo, en esta sección se presentan una serie de conceptos sobre almacenamiento y gestión de datos que son de gran importancia para abordar el diseño de la arquitectura de datos y comunicaciones del Gemelo Digital del Entorno Operativo Marítimo. Dichos conceptos se presentarán de forma progresiva, desde la definición de lo que son datos estructurados y no estructurados y las soluciones tecnológicas para manejarlos, como p. ej. las *Bases de*

Datos (BBDD) NoSQL (Not-only Standard Query Language), hasta los conceptos más avanzados como el de *Data Lakehouse*.

2.4.1 Datos estructurados vs no Estructurados

En [41] se indica que los datos estructurados son aquellos con un formato predefinido y fijo, facilitando su introducción, consulta y análisis. Se componen de campos fijos que permiten una fácil manipulación manual y la identificación de patrones por algoritmos de aprendizaje automático. Ejemplos de datos estructurados son transacciones en línea y datos de supervisión de redes corporativas, etc.

Por otro lado, los datos no estructurados carecen de definición y pueden adoptar diversas formas y tamaños, incluyendo texto, imágenes, audio y video. Representan aproximadamente el 80% de los datos de una organización y son más difíciles de analizar debido a su falta de un formato predefinido. Ejemplos incluyen archivos de medios como podcasts y vídeos, cuyo contenido no puede ser fácilmente buscado sin metadatos adicionales.

Sin entrar en demasiado detalle, tal y como se puede leer en [42] sobre las diferencias entre datos estructurados y no estructurados, los primeros se suelen gestionar en bases de datos relacionales, mientras que los datos no estructurados no se ajustan fácilmente al modelo relacional y son más apropiados para su almacenamiento en bases de datos *NoSQL*, que se tratan en el siguiente apartado. Volviendo a las diferencias entre estos dos tipos de datos, estas son principalmente la flexibilidad, el almacenamiento, el tipo de contenido, la facilidad de análisis y la clasificación. Los datos estructurados son menos flexibles, pero más fáciles de analizar debido a su uniformidad, mientras que los datos no estructurados son más flexibles, pero requieren métodos no convencionales de gestión y análisis.

Entre los datos estructurados y no estructurados, se habla también de un tipo denominado semiestructurados que tendrían cierto nivel de estructura y organización, pero no se ajustan al esquema de datos fijo de los datos estructurados tradicionales [43].

2.4.2 Bases de datos SQL y NoSQL

Este apartado trata la distinción entre bases de datos *SQL (Standard Query Language)* y *NoSQL*, destacando las características y ventajas de cada tipo ([44] y [45]).

En el caso de las bases de datos *SQL*, se trata de bases de datos relacionales que utilizan el lenguaje de consulta estructurado *SQL*. Estas bases de datos poseen una relación predefinida entre elementos, permitiendo la identificación inequívoca de cada registro. Como ventajas se destacan las herramientas disponibles para evitar la duplicidad de registros, un sólido soporte debido a su larga presencia en el mercado, atomicidad de la información y un sistema estándar bien definido para operaciones con la base de datos.

En contraste con las anteriores, las bases de datos *NoSQL* son no relacionales, organizando la información generalmente en documentos y prescindiendo de un identificador para relacionar conjuntos de datos. Se destacan cuatro tipos principales de bases de datos *NoSQL*: valor-clave, de documentos, de grafos y columna ancha. Las ventajas de las bases de datos *NoSQL* incluyen su versatilidad, compatibilidad con una escalabilidad horizontal, disponibilidad, capacidad, y la posibilidad de realizar consultas utilizando lenguajes de *API*, como por ejemplo *JSON*.

La *Figura 2-14* muestra de forma gráfica e intuitiva lo que sería una *BBDD SQL* y *NoSQL*. Como se puede ver, el segundo tipo permite almacenar información más heterogénea.

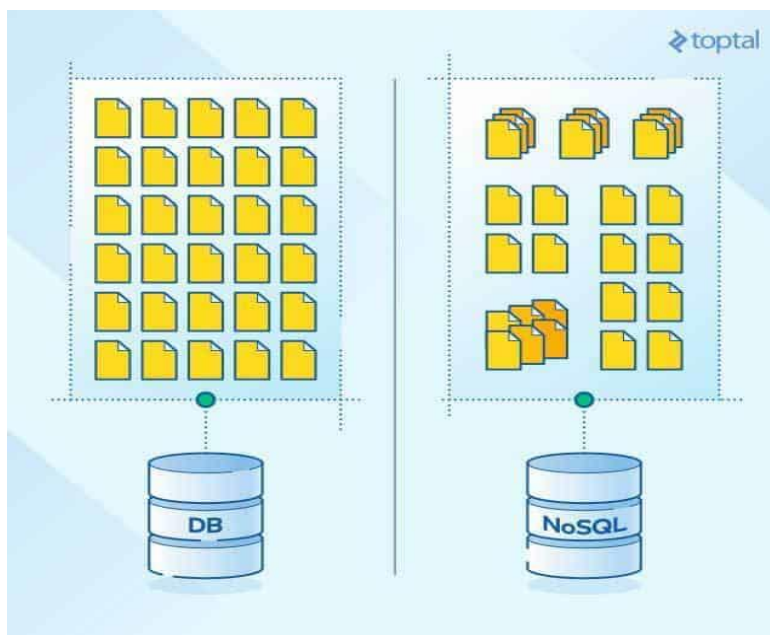


Figura 2-14: Comparativa de BBDD SQL y NoSQL [46]

Mohammad Altrade [47] proporciona un artículo muy completo que compara las *BBDD SQL* y *NoSQL* abordando aspectos como rendimiento, fiabilidad, disponibilidad, consistencia, almacenamiento y escalabilidad. De forma resumida, se indica que las bases de datos *NoSQL* ofrecen un rendimiento superior, son más versátiles en términos de flexibilidad y capacidad de almacenamiento, permiten escalabilidad horizontal y facilitan la programación a los desarrolladores. No obstante, las bases de datos *SQL* son consideradas más fiables y consistentes, adecuadas para conjuntos de datos no extremadamente grandes.

En dicho artículo, se señala que la evolución de *NoSQL* está muy relacionada con el reconocimiento del problema de escalabilidad en *SQL* por parte de empresas como *Google*, *Amazon* y *Facebook*. Esto condujo al desarrollo de sistemas *NoSQL*, tanto propietarias como de código abierto, para abordar las necesidades específicas de grandes empresas con grandes volúmenes de datos e infraestructura. Entre los sistemas más populares se encuentran *BigTable*, *DynamoDB*, *Hypertable*, *Cassandra*, *MongoDB*, *HBase* y *Redis*.

La *Figura 2-15* incluida en la referencia anterior muestra una tabla resumen comparando los dos modelos de *BBDD*:

Feature	NoSQL Databases	Relational Databases
Performance	High	Low
Reliability	Poor	Good
Availability	Good	Good
Consistency	Poor	Good
Data Storage	Optimized for huge data	Medium sized to large
Scalability	High	High (but more expensive)

Figura 2-15: Comparativa de características de BBDD SQL y NoSQL [47]

De cara a tomar la decisión entre una *BBDD SQL* y una *NoSQL*, la siguiente tabla que se puede encontrar en el sitio web de *Azure* de *Microsoft* [48] proporciona una serie de indicaciones:

	NoSQL o no relacional	SQL o relacional
LA MEJOR OPCIÓN PARA:	• Administrar datos de gran volumen, no relacionados, indeterminados o que cambian rápidamente. • Datos independientes del esquema o esquema dictados por la aplicación. • Aplicaciones en las que el rendimiento y la disponibilidad son más importantes que una coherencia alta. • Aplicaciones siempre activas que dan servicios a usuarios de todo el mundo.	• Administrar datos relacionales con requisitos lógicos y discretos que se puedan identificar con antelación. • Esquema que se debe mantener sincronizado entre la aplicación y la base de datos. • Sistemas heredados creados para estructuras relacionales. • Aplicaciones que requieren transacciones de varias filas o consultas complejas.
ESCENARIOS:	• Aplicaciones móviles. • Análisis en tiempo real. • Administración de contenido. • Personalización. • Aplicaciones de <i>IoT</i>. • Migración de bases de datos.	• Sistemas de contabilidad, finanzas y bancarios. • Sistemas de administración de inventario. • Sistemas de administración de transacciones.
ESCALA:	• Escala los datos horizontalmente mediante el particionamiento entre los servidores.	• Escala los datos verticalmente al aumentar la carga del servidor.
MODELO DE DATOS:	• Tipos de base de datos: bases de datos de pares clave-valor, documentos, en columnas y de grafos. • Almacena los datos en función del tipo de base de datos.	• Tipo de base de datos: tablas de filas, agrupadas en relaciones. • Usa el Lenguaje de consulta estructurado (<i>SQL</i>). • Almacena datos como filas en tablas; datos relacionados almacenados por separado y unidos para consultas complejas.

Tabla 2-12: Comparativa de BBDD Relacionales y NoSQL [48]

En relación a la elección de un tipo u otro, resulta de interés el denominado teorema *CAP*, propuesto por el informático *Eric Brewer*, que guía el diseño de sistemas de datos compartidos en red, influyendo en el desarrollo de sistemas distribuidos y en la elección de bases de datos *NoSQL*. Este teorema, que aparece bien explicado en el artículo web de *Barmanand Kumar* [49], establece que en sistemas distribuidos solo se pueden lograr dos de tres garantías para una base de datos: coherencia, disponibilidad y tolerancia de partición.

En términos sencillos, las características son:

- **Consistencia:** Todos los clientes ven los mismos datos al mismo tiempo, sin importar a qué nodo se conecten. Para lograr esto, cada escritura de datos debe replicarse instantáneamente en todos los nodos antes de considerarse exitosa.
- **Disponibilidad:** Cada nodo que no falla debe devolver una respuesta para todas las solicitudes en un tiempo razonable, incluso si algunos nodos están inactivos.

- **Tolerancia de partición:** El sistema debe continuar funcionando a pesar de pérdida de mensajes o fallo parcial del sistema, recuperándose adecuadamente cuando se restaura la comunicación entre nodos.

El teorema CAP clasifica los sistemas en tres categorías en función de las dos garantías que se pueden cumplir:

- **Base de datos CP (Consistente y tolerante a particiones):** Ofrece coherencia y tolerancia a particiones a expensas de la disponibilidad. Durante una partición, el nodo no consistente se cierra hasta resolverla.
- **Base de datos AP (Disponible y tolerante a particiones):** Brinda disponibilidad y tolerancia a particiones sacrificando la coherencia. Durante una partición, todos los nodos permanecen disponibles, pero algunos pueden devolver una versión más antigua de los datos.
- **Base de datos CA (Consistente y disponible):** Ofrece coherencia y disponibilidad en ausencia de particiones de red. Los servidores de bases de datos de un solo nodo suelen clasificarse como sistemas CA.

En sistemas distribuidos, la tolerancia a particiones es esencial. Las particiones y pérdida de mensajes son inevitables, y los diseñadores deben elegir entre coherencia y disponibilidad en función de las necesidades del sistema. La *Figura 2-16* proporciona una representación gráfica del teorema CAP además de productos representativos de cada categoría.

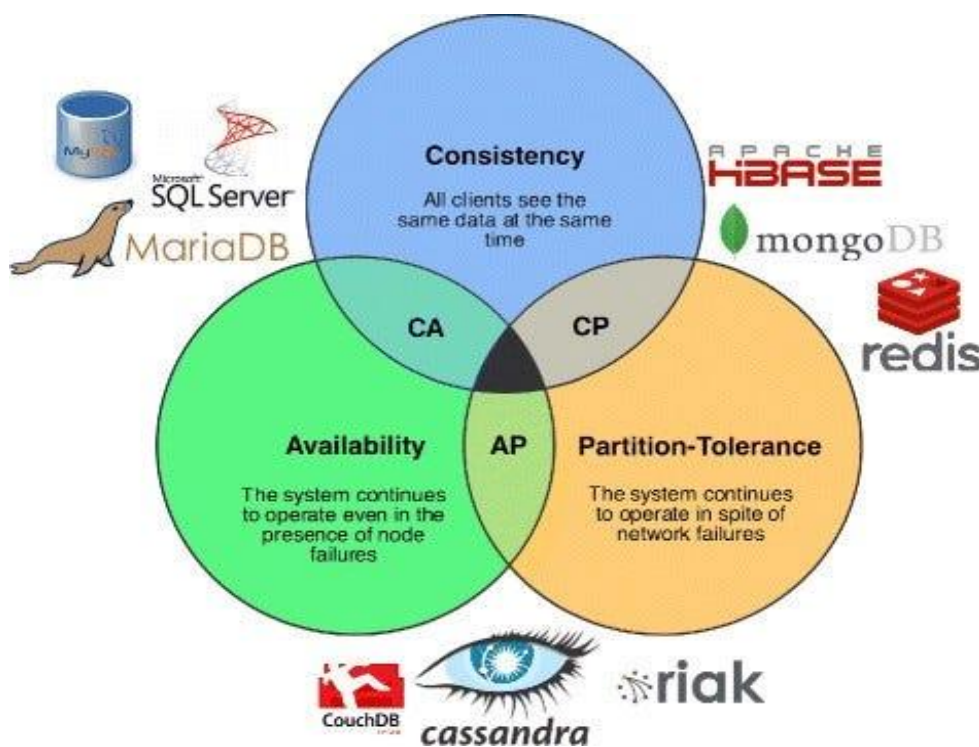


Figura 2-16: Representación gráfica del teorema CAP [49]

Finalmente se muestran en la siguiente tabla las principales características de algunas de las más conocidas alternativas de *BBDD NoSQL* [47].

	<i>Storage Type</i>	<i>Query Method</i>	<i>Interface</i>	<i>Programming Language</i>	<i>Open Source</i>	<i>Replication</i>
<i>Cassandra</i>	<i>Column Store</i>	<i>Thrift API</i>	<i>Thrift</i>	<i>Java</i>	<i>Yes</i>	<i>Async</i>
<i>MongoDB</i>	<i>Document Store</i>	<i>Mongo Query</i>	<i>TCP/IP</i>	<i>C++</i>	<i>Yes</i>	<i>Async</i>
<i>HyperTable</i>	<i>Column Store</i>	<i>HQL</i>	<i>Thrift</i>	<i>Java</i>	<i>Yes</i>	<i>Async</i>
<i>CouchDB</i>	<i>Document Store</i>	<i>MapReduce</i>	<i>REST</i>	<i>Erlang</i>	<i>Yes</i>	<i>Async</i>
<i>BigTable</i>	<i>Column Store</i>	<i>MapReduce</i>	<i>TCP/IP</i>	<i>C++</i>	<i>No</i>	<i>Async</i>
<i>HBase</i>	<i>Column Store</i>	<i>MapReduce</i>	<i>REST</i>	<i>Java</i>	<i>Yes</i>	<i>Async</i>

Tabla 2-13: Principales características de soluciones conocidas de BBDD NoSQL [47]

2.4.3 Big Data

Aunque ya se adelantó en la introducción de esta sección lo que es el *Big Data*, en este apartado se proporciona una descripción más detallada de esta tecnología.

2.4.3.1 ¿Qué es el Big Data?

De forma general, se puede decir que el concepto de *Big Data* se refiere a las problemáticas de almacenamiento, gestión y explotación de cantidades enormes de datos, estructurados y no estructurados, que superan las capacidades tanto del software de procesamiento como la de las arquitecturas de almacenamiento en las *BBDD* tradicionales. Abordar estas problemáticas supone un importante desafío continuo incluso con las tecnologías modernas como el *cloud* o los últimos sistemas de procesamiento y almacenamiento.

Las conocidas "tres V" de *Big Data* -volumen, velocidad y variedad- definen sus características clave [50]. El volumen se refiere al crecimiento exponencial de datos generados, con el 90% de los datos mundiales generados en los últimos dos años. La velocidad destaca la necesidad de procesamiento rápido, especialmente en aplicaciones como transacciones financieras, juegos en línea e *IoT*. La variedad señala la diversidad de formatos de datos más allá de los tradicionales, incluyendo texto, audio, video, etc.

Se mencionan otras "V" como la veracidad, valor y variabilidad, subrayando la importancia de la calidad, valor y diversidad de los datos.

A nivel de cifras, en [50] se destaca que, mientras en el pasado los datos almacenados eran principalmente estructurados, el *Big Data* actual incluye cada vez más datos no estructurados, proyectándose que entre el 80 % y el 90 % de los *Big Data* en 2023 serán no estructurados. Se estima que para 2029 el mercado de análisis *Big Data* alcance más de 655 mil millones de dólares, el doble del valor actual en 2023.

2.4.3.2 Arquitectura y ecosistema Big Data

Una arquitectura de *Big Data* aborda el diseño y organización de hardware y software para manejar conjuntos de datos extremadamente grandes. Se diferencia de la arquitectura de datos tradicional por su capacidad para procesar grandes volúmenes y su enfoque en sistemas de bases de datos distribuidas. Esta arquitectura se centra en datos de cientos de gigabytes, utilizando a menudo sistemas distribuidos para capturar, procesar y analizar datos en tiempo real o casi en tiempo real.

La arquitectura de *Big Data* abarca varias capas fundamentales [50]:

- **Fuentes de Datos:** Se obtienen datos de diversas fuentes en formatos estructurados, semiestructurados y no estructurados.
- **Recopilación e Ingesta de Datos:** Se recopilan datos mediante ingestión por lotes (de forma periódica) o en tiempo real.
- **Almacenamiento de Datos:** Los datos ingeridos se almacenan eficientemente considerando el volumen, tipos de datos, patrones de acceso y requisitos de consulta.
- **Procesamiento de Datos:** Esta capa filtra, combina y representa datos para su análisis. El procesamiento puede ser por lotes o en tiempo real, y los resultados se almacenan típicamente en los denominados *Data Warehouses* o *Data Lakes* que se describen posteriormente.
- **Analítica de Datos:** Involucra análisis tradicionales, búsquedas, análisis predictivos y ad hoc. Se emplea aprendizaje automático para crear modelos y explorar datos sin procesar. Además, se utilizan herramientas de análisis en tiempo real y cuadros de mando de *BI* (*Business Intelligence*).
- **Visualización de Datos:** Se representa información de manera gráfica mediante herramientas como *IBM Cognos*, *Qlik Sense*, *Tableau* y *Power BI*. Facilita la comprensión de datos complejos a las diferentes partes interesadas.
- **Seguridad y Monitorización de Datos:** Implementa protocolos de seguridad, autenticación y cifrado. Monitoriza continuamente ciberataques, *malware* y virus. Gestiona operaciones de gobierno de datos para cumplir con regulaciones. Supervisa el estado y rendimiento de la arquitectura mediante registros, copias de seguridad y alertas del sistema.

Esta arquitectura por capas descrita garantiza la seguridad, disponibilidad y calidad de los datos, permitiendo análisis efectivos y toma de decisiones informada.

En la *Figura 2-17*, disponible en la web de la empresa especializada en analítica de datos SAS [51], se muestran de forma muy gráfica y completa las fases del análisis de datos, desde las fuentes de datos hasta la entrega de valor que coinciden en gran medida con la arquitectura de *Big Data* descrita.

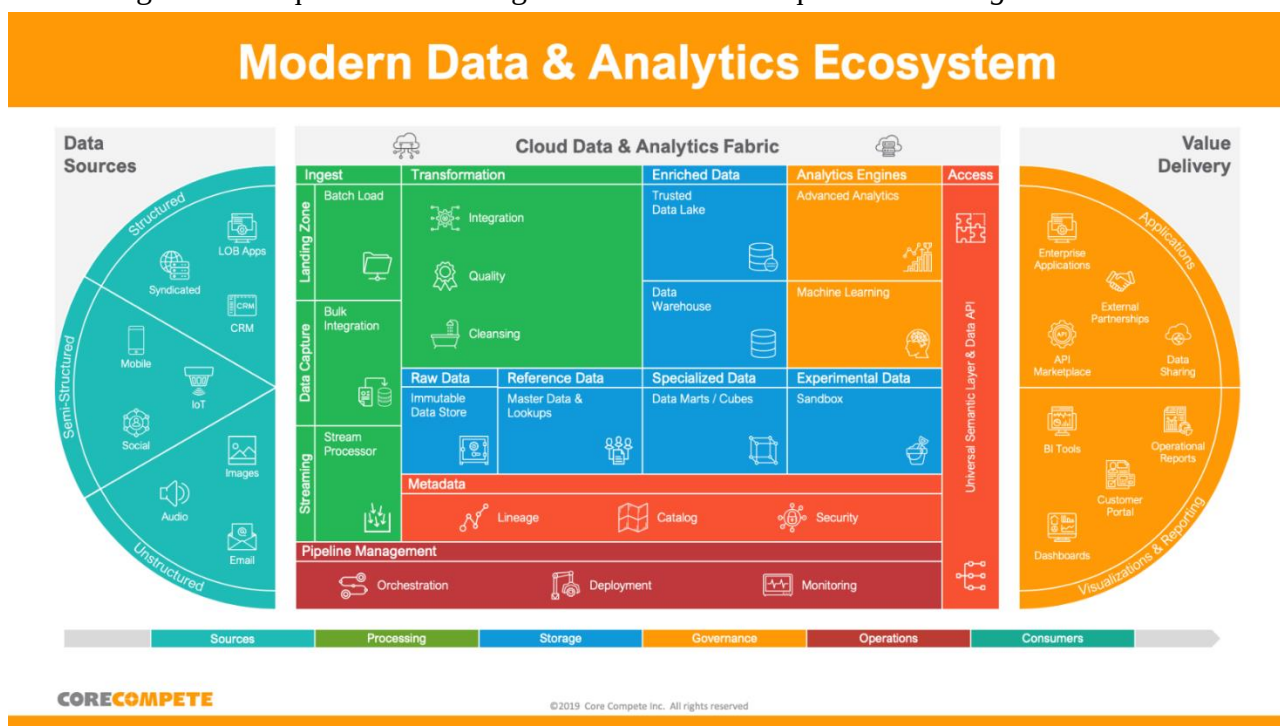


Figura 2-17: Flujo de Analítica de Datos [51]

Respecto a las tecnologías específicas disponibles en el mercado para implementar estas arquitecturas, en el sitio web *Intellspot* [52] se enumeran algunas de las soluciones más utilizadas relacionadas con las tecnologías de Big Data para almacenamiento, minería de datos, análisis y visualización tales como:

Almacenamiento

1. **Apache Hadoop:** Es una destacada tecnología de almacenamiento de *Big Data* que proporciona un ecosistema y marco diseñado para la recopilación, almacenamiento y análisis de grandes conjuntos de datos. *Hadoop* es altamente escalable, distribuye datos automáticamente y puede manejar errores de manera eficiente.
 2. **MongoDB:** Es una base de datos *NoSQL* orientada a documentos como servicio (*DaaS*). Destaca por su accesibilidad en la nube, escalabilidad y flexibilidad. Combina beneficios de bases de datos relacionales con innovaciones de *NoSQL*, permitiendo un desarrollo rápido y almacenamiento de datos en documentos *JSON*.
 3. **Cassandra:** Es una base de datos capaz de manejar grandes cantidades de datos no estructurados. Destaca por su capacidad de escalar rápidamente con tareas administrativas mínimas. Utilizado por grandes empresas como *eBay*, *GitHub* e *Instagram*.
 4. **MarkLogic:** Es una tecnología *NoSQL* empresarial que facilita la integración de datos de silos. Es una plataforma ágil y segura utilizada por grandes organizaciones y pymes para operaciones comerciales críticas, sin necesidad de modelado de datos.
- Riak KV:** Es una base de datos *NoSQL* flexible, optimizada para *IoT* y datos de series temporales. Destaca por su disponibilidad, escalabilidad y facilidad de operación. Ofrece un rendimiento rápido y tolerancia a fallos mediante la distribución automática de datos.

Minería de datos

1. **RapidMiner:** Es una plataforma de ciencia de datos y aprendizaje automático de código abierto. Facilita la creación, entrega y mantenimiento de análisis predictivos, con una interfaz amigable y herramientas integradas.
2. **IBM SPSS Modeler:** Ofrece un conjunto de herramientas dedicadas a la minería de datos, incluyendo preparación de datos, análisis predictivo, análisis de texto y más. Proporciona una interfaz visual para utilizar algoritmos de minería de datos sin necesidad de habilidades de programación.
3. **SAS Enterprise Miner:** Simplifica el proceso de minería de datos para desarrollar modelos rápidamente. Ofrece una interfaz limpia, evaluación de modelos sin codificación y ayuda a identificar *outliers* en los datos.
4. **Weka:** Es un software gratuito de aprendizaje automático desarrollado en la Universidad de Waikato. Ideal para minería de datos y modelado predictivo, con una interfaz gráfica que facilita el acceso a diversas funciones.
5. **Oracle Data Mining (ODM):** Componente de la opción de base de datos de *Oracle Advanced Analytics*, proporciona algoritmos de minería de datos para clasificación, predicción y regresión. Ayuda a predecir el comportamiento del cliente y detectar oportunidades de venta cruzada.

Analítica de datos

1. **Apache Spark:** Herramienta de análisis de *Big Data* de código abierto, fácil de usar y de alto rendimiento para datos por lotes y en tiempo real. Ofrece más de 80 operadores de alto nivel y es compatible con *Python*, *R*, *Scala* y *SQL*.
2. **Apache Kafka:** Plataforma de procesamiento de flujos de código abierto, rápida y escalable. Utilizada para construir tuberías de datos en tiempo real y aplicaciones de transmisión.
3. **IBM Analytics:** Suite de herramientas de análisis empresarial que automatiza procesos de planificación, descubre patrones y factores subyacentes en los datos.
4. **SAP BusinessObjects:** Suite flexible y escalable de herramientas de análisis empresarial que transforma datos en información crucial. Ofrece potentes herramientas de inteligencia empresarial y permite la distribución masiva de información.
5. **Sisense:** Herramientas avanzadas de análisis de datos que permiten a usuarios de negocios y científicos de datos realizar análisis sin necesidad de habilidades de programación. Destaca por sus *dashboards* (cuadros de mando) personalizables y visualizaciones interactivas.

Visualización de datos

1. **Tableau:** Software popular de inteligencia empresarial que facilita la visualización y compartición de datos de manera intuitiva. Conecta con diversas fuentes de datos y proporciona soluciones de análisis únicas.
2. **Looker:** Herramienta de visualización de datos que utiliza *SQL* para organizar datos no estructurados. Ofrece paneles en tiempo real con visuales claros y se conecta con varias bases de datos.
3. **Klipfolio:** Software de tablero de negocios que permite conectar datos de diversas fuentes. Ofrece visualizaciones de datos interactivas y facilita el seguimiento del rendimiento empresarial.

En la *Figura 2-18* se muestran las herramientas destacadas en cada fase o capa de la arquitectura *Big Data*.

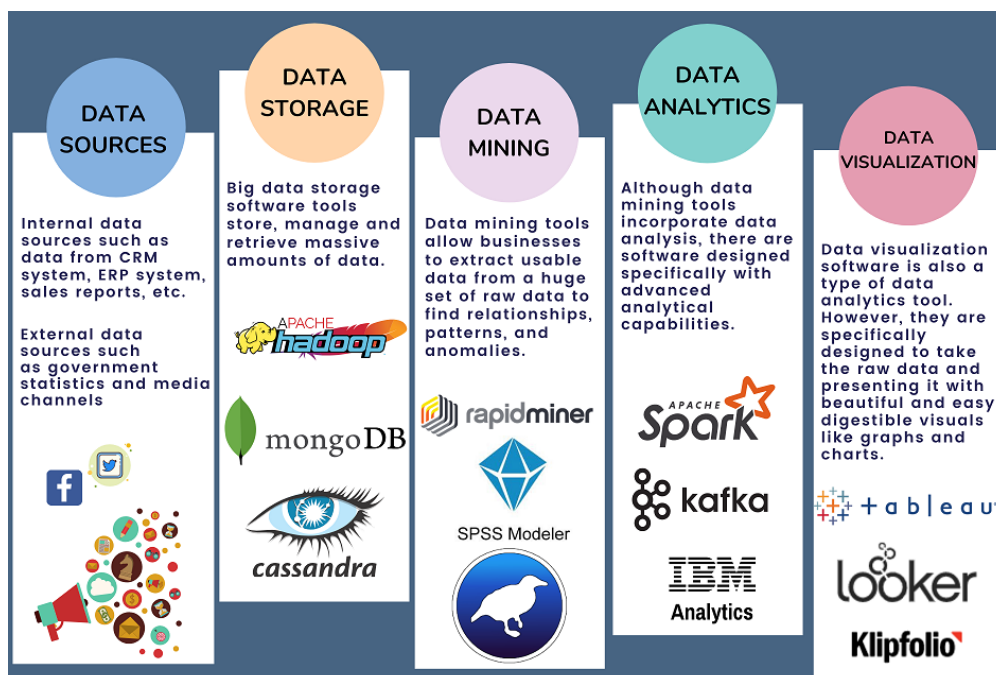


Figura 2-18: Herramientas destacadas para implementación de *Big Data* [52]

2.4.3.3 Data Warehouse, Data Lake y Data Lakehouse

Íntimamente relacionados con el *Big Data* y la analítica de datos, se presentan los conceptos mencionados previamente de Almacenes de Datos (*Data Warehouses*), Lagos de Datos (*Data Lakes*), y el más reciente *Data Lakehouse* que es una especie de combinación de ambos. Obviamente, estos tres conceptos se refieren a arquitecturas y sistemas de almacenamiento de los datos para su posterior utilización. En la referencia [53], *John Kutay* presenta de forma muy completa estos tres conceptos, que se describen de forma resumida a continuación:

El **Data Warehouse** (DW) es una estructura diseñada para organizar grandes cantidades de datos estructurados de diversas fuentes, permitiendo el análisis de alto rendimiento y simultáneo de múltiples usuarios. Al estructurar los datos, mejora la productividad para consultas específicas y su integración en aplicaciones analíticas, como *Business Intelligence* (BI). Sin embargo, su limitación a datos estructurados y su vinculación a software propietario son desafíos importantes. En la siguiente *Figura 2-19* se muestra un esquema de un *Data Warehouse*.

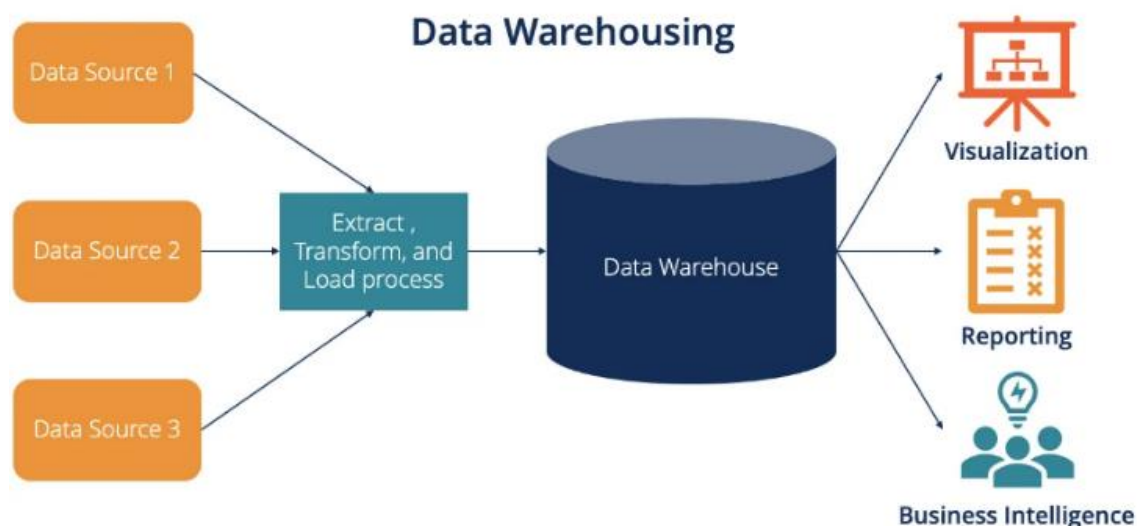


Figura 2-19: Esquema de *Data warehouse*. Fuente <https://corporatefinanceinstitute.com/>

En contraste, el **Data Lake** es un repositorio de datos donde, inicialmente, se almacenan sin procesar y sin un esquema unificado. Esto facilita la disponibilidad para futuros usos, con la posibilidad de procesarlos y traducirlos a un esquema corporativo. Este esquema es ideal para reutilización a largo plazo y por diferentes departamentos de una organización. Además, presenta ventajas como rapidez en la creación y análisis de modelos, costos reducidos de hardware y el uso de tecnologías de código abierto. Las principales ventajas del Data Lake son las siguientes:

1. Almacenamiento de datos en bruto para uso futuro.
2. Rapidez en la creación y análisis de nuevos modelos.
3. Costos reducidos de hardware y uso de tecnologías de código abierto.
4. Reducción del gasto de recursos, principalmente al utilizar datos.

Los principales retos del *Data Lake*:

1. Complejidad en el despliegue y gestión, incluyendo el crecimiento de la base de datos y posibles redundancias.
2. Necesidad de actualizar la base de datos en casos como solicitudes de derecho al olvido.
3. Complementariedad con *Data Warehouse* para obtener valor de los datos.

4. Gestión del histórico de versiones de datos y la necesidad de fusiones, actualizaciones y eliminaciones.
5. Desafíos para personas con menos conocimientos profesionales al analizar datos no estructurados.

La Figura 2-20 muestra un esquema de una *Data Lake*:

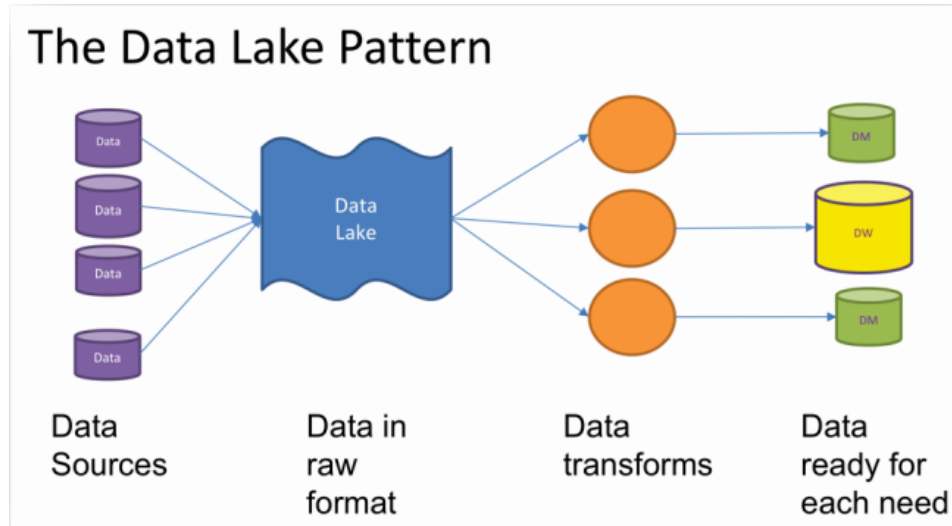


Figura 2-20: Esquema de *Data lake*. Imagen de datakitchen.io

En los últimos años ha surgido el concepto de ***Data Lakehouse***, que combina lo mejor de *Data Lake* y *Data Warehouse*, aprovechando las fortalezas de ambas arquitecturas reuniendo en un único repositorio todos los datos (estructurados, semi-estructurados y no estructurados). Permite la integración de herramientas de código abierto y optimiza la utilización de los datos. Suelen comenzar como Data Lakes que contienen todos los tipos de datos; estos se convierten a un formato Delta Lake, que es una capa de almacenamiento de código abierto que aporta confiabilidad a los Data Lakes permitiendo así procesos transaccionales **ACID** (**Atomicity, Consistency, Isolation, and Durability**). No obstante, aún es una tecnología inmadura por lo que su aplicación implica ciertos riesgos. En la Figura 2-21 se muestra un esquema de un *Data Lakehouse*:

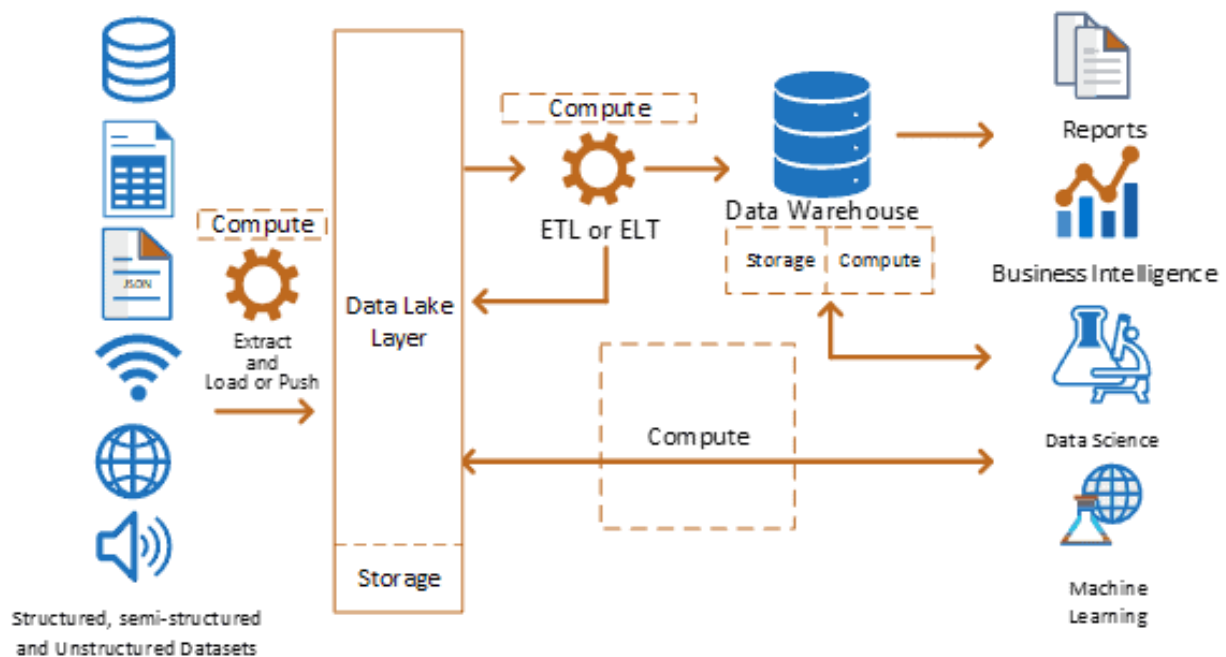


Figura 2-21: Esquema de un *Data Lakehouse* [54]

En resumen, mientras el *Data Warehouse* es ideal para consultas estructuradas, el *Data Lake* ofrece flexibilidad y almacenamiento en bruto para un uso más amplio y diverso de los datos. La combinación de ambos en un *Lakehouse* permite abordar de manera integral las necesidades de almacenamiento y análisis de datos en entornos empresariales complejos.

2.5 Sistemas de comunicaciones en el entorno naval

El objetivo de esta sección es presentar las capacidades de los sistemas de comunicaciones para plataformas navales actuales en el sector de la Defensa y que, por la propia naturaleza de los escenarios en los que desempeñan sus misiones, suponen un reto importante para la implementación y operación de un gemelo digital, condicionando en gran medida la arquitectura del sistema y la forma de operar. Por tanto, se ha hecho una pequeña investigación, entre dos de los principales fabricantes del sector, de las soluciones que ofertan y sus principales características, que ya de por sí se enfrentan a los retos de cumplir con todos los requisitos de las operaciones navales que requieren comunicaciones seguras y de alta velocidad, tanto entre buques desplegados y fuerzas en teatros remotos, como con los centros de mando e incluso con las familias en casa.

Las dos tecnologías más importantes para las comunicaciones a larga distancia de las plataformas navales desplegadas son los sistemas de comunicaciones vía satélite y las comunicaciones radio *HF* (*High Frequency*).

Según se puede leer en su sitio web [55] y [56], Thales ha diseñado estaciones y terminales *satcom* específicamente para buques de superficie y submarinos, asegurando acceso rápido, alta disponibilidad, uso eficiente de recursos satelitales y protección contra ciber amenazas. Estas estaciones integran un sistema de estabilización de tres ejes para apuntar directamente al satélite, resisten condiciones militares y se adaptan a las limitaciones submarinas, incluyendo la presión. Además, se garantiza la compatibilidad entre las comunicaciones satelitales y otros sistemas a bordo.

Por ejemplo, en la compañía Thales, la terminal *DiveSat* está diseñada para submarinos, mientras que la familia *SurfSat* está diseñada para barcos de superficie (*SurfSat L* para barcos más grandes, *SurfSat S* para embarcaciones más pequeñas). La *Figura 2-22* muestra gráficamente el alcance de las soluciones de comunicaciones satélite de *Thales* para las plataformas navales.



Figura 2-22: Imagen promocional del alcance de las soluciones *satcom* navales de *Thales* [56]

Entre las principales características de las soluciones proporcionadas destacan:

- Operación en todo el mundo: seguimiento automático por satélite incluso en elevaciones altas y bajas, sin zonas ciegas.
- Funciones multibanda (X , Ka)/antena dual/ funciones de gestión (*Handover*) de satélite dual.
- Terminales independientes del módem. Con *Modem21*, total compatibilidad con Guerra Electrónica y *anti jamming*.
- Todos los niveles de servicio disponibles: rendimiento que oscila entre 256 kb/s y 20 Mb/s.

Respecto a las comunicaciones mediante radio HF, se destaca su papel crucial en las comunicaciones tácticas y estratégicas militares, como un recurso alternativo a la comunicación por satélite. Las evoluciones recientes, como el establecimiento automático de enlaces y sistemas de tercera generación, facilitan la configuración y mantenimiento de enlaces *HF*. La llegada de la banda ancha *HF* ofrece velocidades de datos significativas con alcances notables. Futuras mejoras prometen mayor confiabilidad y robustez en la comunicación *HF*, adaptándose automáticamente a las condiciones cambiantes del canal.

Thales aborda los desafíos de enlaces deficientes e interferencias con tecnologías como formas de onda de banda ancha y *HFXL*, permite configuraciones remotas, proporciona velocidades de datos muy altas con formas de onda *4G HF*, facilitando la integración en redes *IP* (*Internet Protocol*) y ampliando su uso en servicios de largo alcance, todo ello cumpliendo con estándares militares, asegurando la interoperabilidad y con un importante enfoque en la ciberseguridad.

Como solución principal de la compañía en esta categoría destaca *AQUILON*, una solución fiable que combina arquitectura *IP* abierta, modular y escalable con seguridad sólida para cumplir con los requisitos clave de comunicaciones de las fuerzas navales [57]. Destaca por su facilidad de integración, uso y mantenimiento, reduciendo la carga de la tripulación. Ofrece capacidad total en todo el espectro de frecuencias, gestión centralizada, nueva generación de componentes y desarrollo en colaboración con usuarios operativos.

En la *Figura 2-23* se da una visión general de los diferentes sistemas de comunicaciones integrados en la arquitectura *IP* de *Thales*.



Figura 2-23: Representación gráfica de la arquitectura *IP* de comunicaciones navales de *Thales* [57]

La compañía Leonardo, por su parte, hace referencia a un concepto denominado *MDA* (Conciencia del Dominio Marítimo) [58]. Este concepto implica conciencia de amenazas y situacional, fusionando información operativa, de inteligencia y medioambiental. En este sentido, las comunicaciones navales modernas se enfrentan a los desafíos emergentes de la *MDA* y se equipan con centros de mando avanzados, integrando redes y recursos radio *HF*, *VHF/UHF* y satélite.

Además, esa creciente importancia de la conciencia del dominio marítimo impulsa la demanda de comunicaciones estandarizadas de voz, video y datos, así como redes móviles ad hoc multiservicio de alta velocidad. La necesidad de radios definidas por software (*SDR*) es evidente, y la transmisión de video y videoconferencias se vuelve esencial para operaciones específicas como la lucha contra la piratería.

La interoperabilidad se hace vital para operaciones conjuntas, destacando la importancia de la comunicación multidominio, incluyendo barco-barco, costa-barco-costa, aire-barco-aire y satélite-barco-satélite. Las comunicaciones internas y externas deben garantizar un intercambio de información fiable, resistente a interferencias y cifrado, especialmente en acciones militares.

En las Figuras 2-24 y 2-25 se puede ver de forma muy resumida el espectro de soluciones de comunicaciones navales ofrecidos por Leonardo, que incluyen tanto las comunicaciones radio en su gama de productos *SWave*, con cobertura en prácticamente todas las bandas (*LF*, *MF*, *HF*, *VHF* y *UHF*), como las comunicaciones vía satélite. Además, incluyen servicios de *VoIP* (Voice over IP), *Link-11,16* y 22, mensajería (*MHS*), etc.

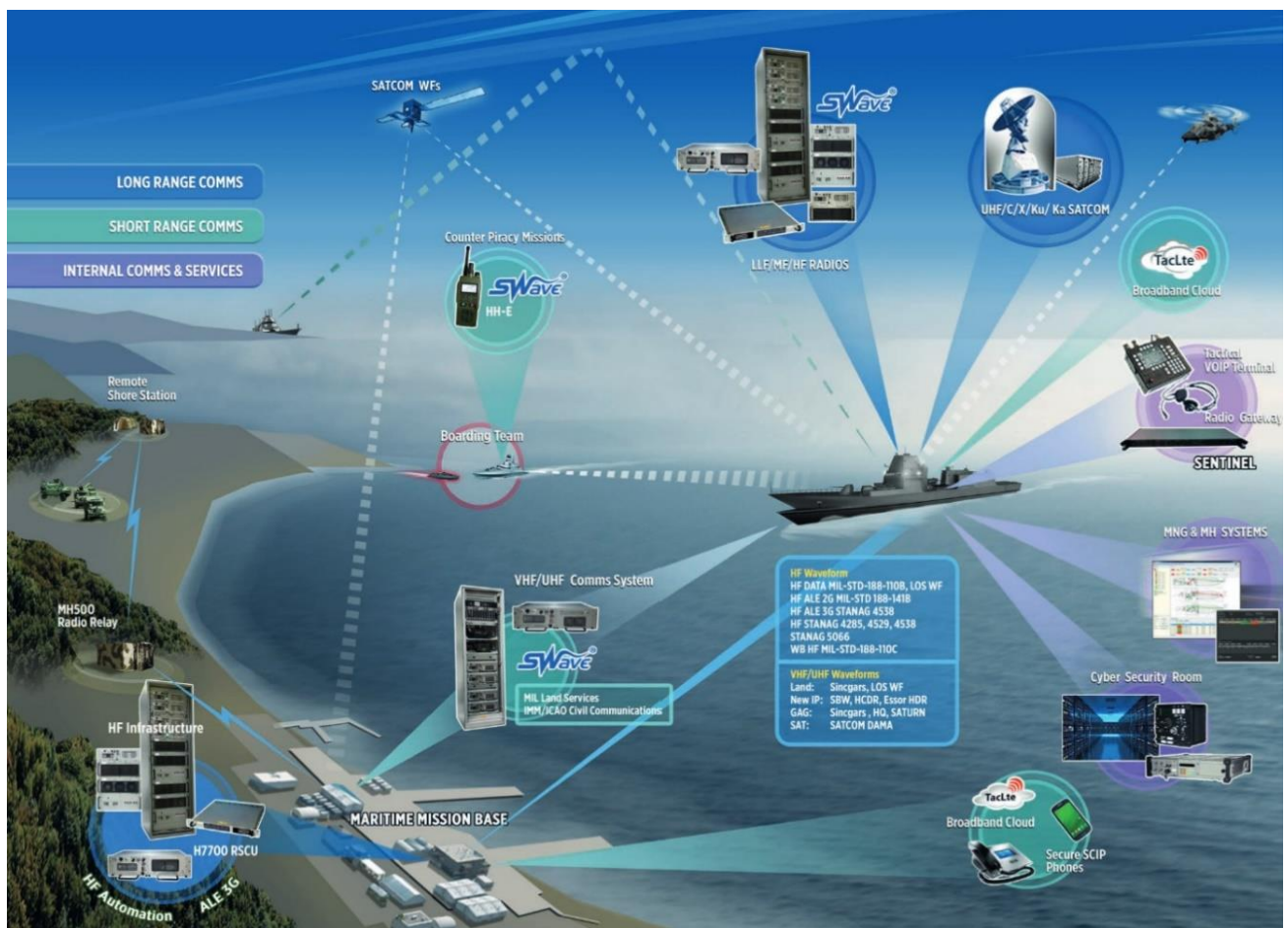


Figura 2-24: Diferentes soluciones de comunicaciones navales de Leonardo [58]

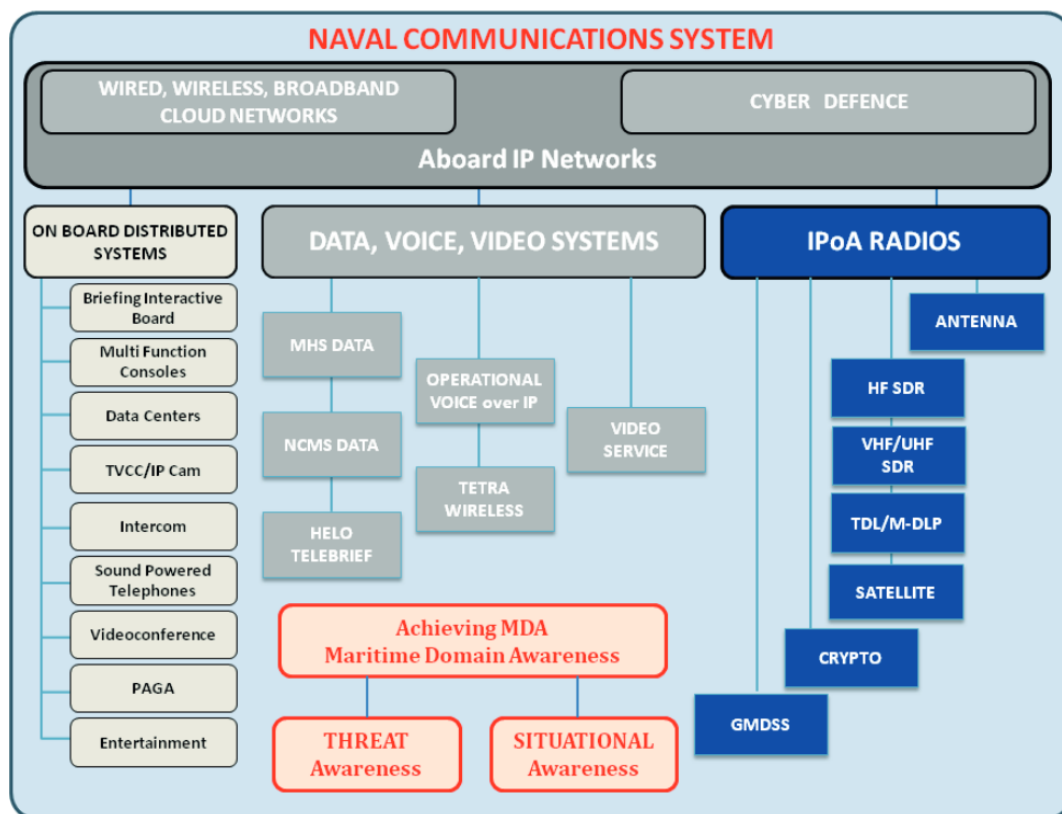


Figura 2-25: Diagrama de bloques de los sistemas de comunicaciones navales de *Leonardo* [58]

Como se puede ver, la evolución de las comunicaciones navales se alinea con la creciente importancia de la conciencia del dominio marítimo. La introducción de subsistemas avanzados y la adopción de tecnologías como *SDR* y *LTE* reflejan el compromiso con operaciones seguras, eficientes e interoperables en un entorno marítimo complejo.

2.6 Sistemas de cuadrícula jerárquica para modelado del entorno

En esta última sección del capítulo 2, se ha querido introducir como concepto teórico las alternativas de sistemas de modelado de información geoespacial mediante cuadrículas o rejillas jerárquicas. La razón de incluir este concepto es que, al final, tanto para la integración como para el análisis y explotación de los datos geoespaciales, se debe definir o disponer de una estructura efectiva y eficiente que proporcione la información necesaria con la mínima sobrecarga y la mayor sencillez posibles, más aún en un escenario en el que existen restricciones operativas en cuanto a comunicaciones y procesamiento de datos. En este sentido, los sistemas de cuadrículas jerárquicos se presentan como opciones bastante adecuadas por su capacidad de manejar en una misma representación diferentes resoluciones con un diseño conceptual sencillo y que permite mantener controlada la carga de procesamiento y tamaño de la información.

Por tanto, tal y como se indica en la web de *Spectus* [59], los sistemas de cuadrícula son fundamentales para analizar grandes conjuntos de datos espaciales, ya que dividen de forma efectiva áreas de la Tierra en celdas de cuadrícula identificables. Utilizando diversas metodologías de cuadrícula estándar de la industria, el análisis de una o varias fuentes de datos puede orientarse con mayor precisión y normalizarse más fácilmente.

Algunos de los sistemas de cuadrícula más utilizados son *Geohash*, *H3*, *S2* y *Bing*. Cada uno de estos, con sus propias peculiaridades, permite a los equipos de datos identificar con precisión áreas de



Figura 2-27: Ejemplo de cuadrícula S2 [549]

2.6.3 Teselas de Bing

Bing Maps proporciona un mapa mundial que utiliza su propio sistema de cuadrícula basado en una representación de mapa plano utilizando la proyección de *Mercator* como base para su sistema de cuadrícula jerárquico, con el problema de la distorsión de escala cuanto más se aleja del ecuador. *Bing* utiliza un sistema de cuatro claves; las coordenadas *XY* del mosaico bidimensional se combinan en cadenas unidimensionales. El número de caracteres de una clave cuádruple es igual al nivel de detalle del mosaico correspondiente. En *Figura 2-28* se puede ver la división del mapa de la Tierra según este esquema.

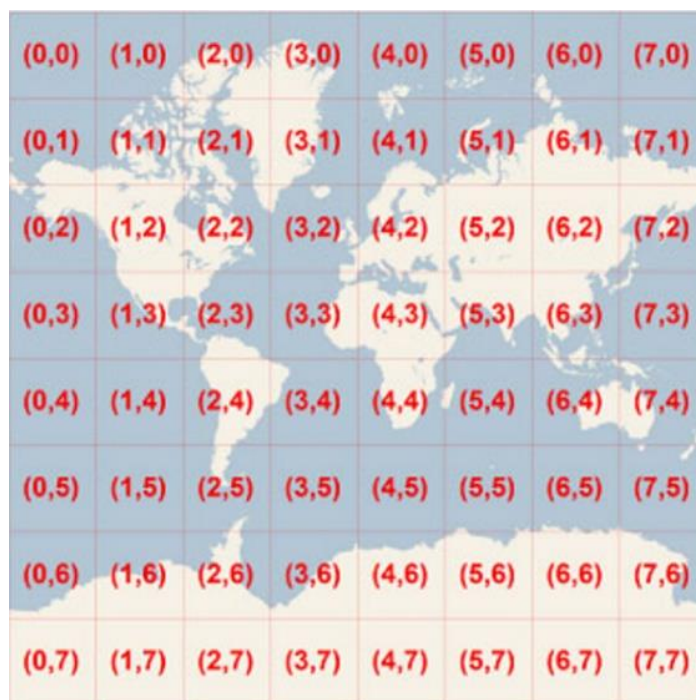


Figura 2-28: Ejemplo de teselas Bing [59]

2.6.4 H3 de Uber

El sistema de red *H3* fue diseñado por *Uber*. Basándose en su modelo de negocio y la necesidad de optimizar los viajes compartidos, desarrollaron *H3* para visualizar datos espaciales. Otros analistas han adoptado el enfoque hexagonal del sistema de cuadrícula para un análisis más preciso de grandes conjuntos de datos espaciales [59].

En la página oficial de *H3* [60] se detallan en profundidad las características y consideraciones técnicas de este sistema. Ahí se indica que es una herramienta de indexación geoespacial que divide el mundo en celdas hexagonales, y desarrollado para abordar las necesidades de ciencia de datos de *Uber*. *H3* destaca fundamentalmente por:

- proporcionar una solución eficiente y flexible para el manejo de datos geoespaciales;
- rellenar de forma óptima el espacio, con un margen de error promedio menor que con mosaicos cuadrados;
- constituir un índice geoespacial jerárquico que permite mapear la Tierra esférica en 16 niveles asignando índices a todas las celdas según la jerarquía espacial;
- ser útil para unir y representar conjuntos de datos dispares con diferentes resoluciones; y
- permitir aplicar aprendizaje automático a datos geoespaciales.

En la *Figura 2-29* se muestra un ejemplo de división de un área geográfica en celdas *H3* combinando diferentes resoluciones.

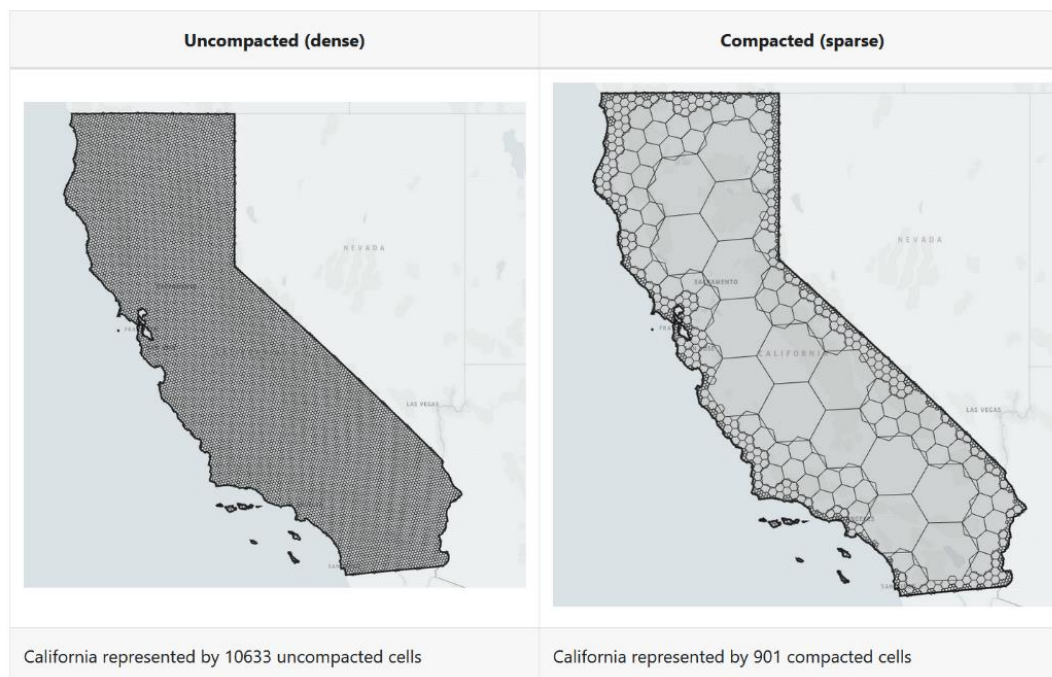


Figura 2-29: Ejemplo de representación *H3* de parte de California [5560]

Debido a las características de este sistema de representación, se considera el más apropiado como base para el desarrollo del gemelo digital. García Velera y Otros [51] utilizan este modelo para caracterizar un entorno marino en una implementación de gemelo digital oceánico. No obstante, tiene la desventaja de que es un sistema de mapeo 2D, que para muchas variables de estudio será adecuado, pero para otras variables podrá requerir complementarlo con la dimensión de altitud o profundidad.

3 ARQUITECTURA DEL SISTEMA

En este capítulo se aborda la propuesta de diseño conceptual de arquitectura del Gemelo Digital de Entorno Operativo Marítimo. A modo de recapitulación, la idea de este gemelo digital es la de dotar a una plataforma naval de un modelo predictivo fiel del entorno marítimo en el que opera y que proporcione información precisa sobre su evolución, mejorando así la consciencia situacional, lo que se podrá traducir en una mayor seguridad, eficacia y eficiencia en el desarrollo de la misión. Este modelo del entorno que se ejecuta en la plataforma, no es el gemelo digital propiamente, sino una especie de modelo-copia del “verdadero” gemelo digital que corre en un *CPD* terrestre. No obstante, este modelo-copia de la plataforma también evolucionará de forma local alimentándose tanto de datos remotos proporcionados por el servicio terrestre, bajo petición de la plataforma, como de datos locales obtenidos por la propia plataforma y otras plataformas auxiliares, u otras de la misma naturaleza que operen en las proximidades. Pero, además, los datos capturados por la plataforma, también se podrán enviar al sistema terrestre en diferentes situaciones y diferentes formatos dependiendo, fundamentalmente, de la conectividad disponible.

En las siguientes secciones se detallarán los pormenores de esta solución, empezando por la arquitectura global del sistema, para luego profundizar en cada uno de los elementos que forman parte y su papel en este sistema de Gemelo Digital de Entorno Operativo Marítimo.

3.1 Arquitectura global

A lo largo del trabajo ya se han ido dando pinceladas de la arquitectura del sistema de Gemelo Digital de Entorno Operativo Marítimo propuesta basada en tres niveles, donde el primer nivel corresponde al sistema central terrestre, con gran potencia de cómputo y comunicaciones, que crea y mantiene el gemelo digital usando todas las fuentes de datos disponibles. El segundo nivel corresponde a la plataforma naval, que desplegará en escenarios operativos con comunicaciones limitadas y ejecutará un modelo local del gemelo digital actualizado con datos propios y datos proporcionados por el sistema central bajo petición. En el tercer nivel se situarían las plataformas desplegables desde la plataforma naval, que proporcionarán información adicional a esta última. No obstante, es en esta sección, donde se va a entrar en detalle en la arquitectura del sistema, primero desde un nivel conceptual de bloques o elementos que forman parte del mismo y sus capacidades, y posteriormente desde un nivel más específico referido a la integración de los datos e información entre esos diferentes componentes del sistema.

Una cosa que se quiere señalar es que la solución propuesta se basa en un entorno “*on-premise*” o, en todo caso, de nube privada, esto es, no se ha planteado recurrir a tecnologías de nube pública como *AWS* o *Azure* para la implementación del sistema. La principal razón reside en la palabra “operativo” incluida en el título del *TFM*, y que de forma implícita se refiere a un escenario relacionado con la Defensa, donde la nube pública tiene difícil cabida por las implicaciones de seguridad que conlleva. No obstante, en un entorno sin estas restricciones, sería perfectamente implementable el modelo en una arquitectura de nube pública, aunque quizás los costes que implicaría no lo harían muy viable.

Como una imagen vale más que mil palabras, se incluye la *Figura 3-1* que proporciona una representación gráfica del sistema completo.

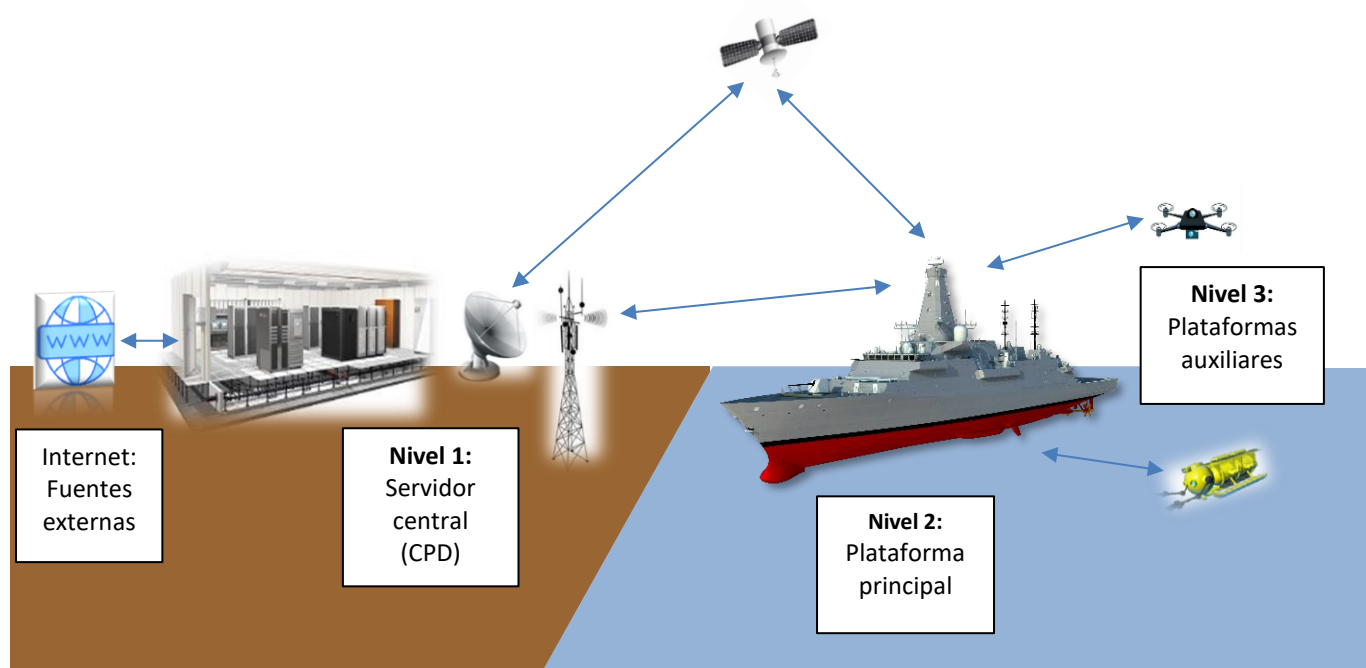


Figura 3-1: Componentes del sistema del Gemelo Digital de Entorno Operativo Marítimo [Elaboración propia con imágenes obtenidas de *depositphotos.com*]

En la figura se muestran los tres niveles mencionados, así como los flujos de comunicación entre ellos, incluyendo el satélite, que proporcionará cobertura donde otros medios no llegan. Teniendo en mente esta imagen de la arquitectura de alto nivel y con una idea más o menos clara de la operativa general del sistema, se pasa a describir en las siguientes secciones cada uno de los niveles de forma detallada.

3.2 Nivel 1: Servidor central en tierra

Como ya se ha mencionado, este es el nivel donde se desarrolla el gemelo digital. El desarrollo consistirá básicamente en el típico ciclo de *Big Data* que se mencionó en uno de los apartados del Capítulo 2 compuesto de una serie de fases que se detallan a continuación, en este caso ya específicamente desarrolladas en detalle para el sistema propuesto:

- **Recopilación de datos:** consistirá en recopilar todos los datos relevantes de todas las fuentes que se determinen, en este caso tanto fuentes institucionales como fuentes propias de datos obtenidos de las plataformas. Como se ha visto, esta información procedente de múltiples fuentes no tendrá un formato homogéneo, habrá datos más estructurados, otros semiestructurados y otros directamente no estructurados: datos como registros en binario o texto, en formato *JSON*, o *GML*, o empaquetados conforme *NetCDF* o *shapefile*, imágenes, videos, series temporales, formatos vectoriales, formatos ráster, etc. Los datos geoespaciales manejarán diferentes resoluciones y cubrirán diferentes áreas geográficas. La generación de nuevos datos e información será continua, con diferentes frecuencias, con predicciones para diferentes periodos de tiempo. Y para complicarlo aún más, el acceso a los diferentes datos requerirá el uso de diferentes *APIs* que además irán evolucionando. Como se puede ver, el acceso y descarga de los datos de las diferentes fuentes supone un desafío enorme. En este sentido, los proyectos que se han mencionado sobre Gemelos Digitales Oceánicos (*DTO*) y fuentes institucionales enfocados en

el desarrollo de sistemas de acceso unificado a los datos de múltiples fuentes supondría un importante avance en el desarrollo de sistemas como este.

- **Almacenamiento de Datos:** al igual que la parte de acceso y descarga de datos, que ya supone un desafío importante por el volumen, velocidad y variabilidad (las 3 “V” de *Big Data*) y forma de acceso a los datos, el almacenamiento va a verse directamente afectado por toda esta misma problemática. Hay que partir del supuesto de que el crueso de los datos descargados son relevantes para el desarrollo de un gemelo digital, que la tecnología para la elaboración de modelos *data-driven* requieren la disponibilidad de esos datos, y si es posible en formatos crudos, ya que determinadas transformaciones podrían alterar determinados patrones de información implícitos en los datos que podrían impedir su correcta detección e identificación por las tecnologías de modelado. Además, se debe asumir que estas tecnologías van a evolucionar y seguramente se requiera reutilizar todos los datos históricos para el desarrollo de los nuevos modelos. El resultado de esto es que, en principio, el sistema debe ser capaz de almacenar todos esos datos que se descargaron y que se seguirán descargando de las diferentes fuentes de forma indefinida y en sus formatos originales o con transformaciones que no alteren su naturaleza o impliquen una pérdida de información.

De lo anterior, y teniendo en cuenta lo que se presentó en la sección que trataba sobre los sistemas de almacenamiento y gestión de datos en el capítulo 2, parece que la opción más adecuada como sistema de almacenamiento sería una infraestructura distribuida sobre la que instalar una base de datos *NoSQL* de tipo documental. Este tipo de *BBDD* proporcionan una gran flexibilidad que permite manejar cualquier tipo de datos y enriquecerlos con los metadatos que se consideren necesarios o útiles. Pero no solo la flexibilidad determina que este tipo de *BBDD* son más adecuadas para esta aplicación; la escalabilidad horizontal y el rendimiento también constituyen factores clave por el volumen de datos y su crecimiento esperado. Además, se debe tener en cuenta que la naturaleza del sistema no implica como factor crítico la consistencia, que es la característica de la que adolecen principalmente las *BBDD NoSQL*, pero en este caso los datos descargados en principio son datos de consulta que no se modificarán una vez almacenados y no parece muy probable que diferentes usuarios vean versiones distintas del mismo dato. Dicho de otra forma, la lógica de negocio del sistema no tiene un carácter transaccional, sino más bien consultivo.

Por tanto, el sistema de almacenamiento que se considera más adecuado será un sistema distribuido con una *BBDD NoSQL*, sobre la que se descargan los datos de las diferentes fuentes. Este sistema de almacenamiento constituirá lo que se denomina un *Data Lake*, ya que las características identificadas se corresponden con la descripción de este tipo de arquitecturas de almacenamiento presentadas en el capítulo anterior, sobre todo en lo referente a su idoneidad para almacenar datos no procesados y con formatos no uniformes.

- **Procesamiento de Datos:** Esta capa forma parte del típico ciclo *ETL* (*Extract, Transform and Load*), como capa de transformación de los datos para prepararlos para su posterior utilización. Los datos en bruto descargados en el sistema requerirán aplicar ciertos procesos para convertirlos en formatos útiles para las herramientas de análisis o explotación de dichos datos. Entre estos procesos se encuentra la depuración de datos, que puede consistir en filtrar valores incorrectos, o determinar qué hacer con los datos faltantes. También puede incluirse en este proceso un enriquecimiento de los datos agregando metadatos, o conversiones de formatos a uno común,

por ejemplo, convertir a *JSON* los paquetes *shapefile*, o a más bajo nivel transformar cadenas en valores numéricos o vice versa. También podría hacerse en esta etapa la transformación de datos *GIS* con diferentes resoluciones y formatos en rejillas *H3* de *Uber* completando datos faltantes mediante interpolaciones, diezmados, copias, medias, etc. En definitiva, puede haber un sinnúmero de posibles operaciones de transformación que se puedan necesitar hacer en esta etapa para proporcionar unos datos más adecuados a las siguientes etapas. De hecho, la forma de usar los datos o las aplicaciones que los ingieran condicionarán en gran medida los requisitos de transformación de los datos. En función de estos usos, los datos transformados se pueden volcar en nuevos *Data Lakes* en formatos no uniformes, o bien en *Data Warehouses* en formatos relacionales para su uso por herramientas de analítica de datos basadas en *SQL*, por poner algunos ejemplos.

- **Analítica de Datos:** esta fase, aunque se ha denominado analítica, incluye todos aquellos procesos que utilizan los datos para obtener valor de los mismos. Dependiendo de la aplicación y analítica, este valor significará reducir la incertidumbre en la toma de decisiones, encontrar determinada información, detectar patrones de amenazas en un análisis de monitorización de redes, o detectar problemas o patrones de uso que permitan optimizar recursos, etc.

En el caso que aplica a este *TFM*, los datos se van a utilizar para desarrollar un modelo *data-driven* usando técnicas de Inteligencia Artificial en sus diferentes variantes, aprendizaje automático, ya sea profundo o no, métodos de regresión, métodos de aprendizaje supervisado y no supervisado, etc. En este punto se quiere recalcar que no es el objetivo de este trabajo profundizar en la parte de modelado del sistema, simplemente se asumirá que, a partir de los datos que se han descargado y procesado, se entrenarán uno o varios modelos mediante técnicas de *IA* que proporcionarán toda aquella información y modelos que puedan resultar de utilidad o formar parte del gemelo digital del entorno operativo. Este gemelo digital será capaz de proporcionar estimaciones fiables de los estados actuales, así como predicciones futuras de las variables que afectan al entorno marítimo y que se consideren relevantes para el funcionamiento del mismo y las operaciones. Además, también será capaz de integrar nueva información recibida periódicamente para mejorar sus estimaciones sin necesidad de un reentrenamiento de todo el modelo. No obstante, debido a la rápida evolución de la *IA*, se debe contemplar la aparición de nuevas técnicas de modelado que requieran un reentrenamiento completo del modelo, como ya se indicó previamente.

Este modelo “vivo” al que se denomina gemelo digital de entorno marítimo funcionará como un servicio con dos propósitos dentro del sistema completo:

1. ejecutándose en el sistema central se irá actualizando con los nuevos datos que reciba y, al mismo tiempo,
2. podrá ser requerido vía servicio remoto para que proporcione un informe o modelo de condiciones de navegación en una zona determinada y durante un periodo de tiempo indicado.

En este último propósito será requerido normalmente por las plataformas para actualizar sus modelos a las condiciones actuales, obviamente, cuando puedan acceder al servicio.

Por otro lado, el modelo entrenado, o más probablemente un submodelo entrenado en el sistema central, también podrá ser descargado en una plataforma naval en algún momento en el

que tenga conectividad adecuada con el sistema central (p.ej. en puerto antes de embarcarse en una misión) para ejecutarse sobre la propia plataforma y generar información y predicciones actualizadas cuando esta se encuentre desplegada y no disponga de conectividad con el sistema central. En este caso se alimentará, por un lado, con unos datos iniciales del escenario que se hayan solicitado y descargado vía servicio remoto en algún momento y, por otro lado, con los datos locales que la plataforma vaya produciendo en su operación y que se espera que contribuirán a mejorar las predicciones.

El funcionamiento de estos servicios de actualización de datos y del modelo que se ejecuta en la plataforma se tratan más en detalle en el siguiente capítulo sobre los casos de uso operativos.

- **Visualización de Datos:** Esta fase asociada al ciclo de *Big Data*, si bien aporta información relevante para el seguimiento, control y comunicación del proyecto y requiere un esfuerzo en su diseño, no tiene implicaciones especialmente claves en el diseño y operatividad del sistema. Los productos de esta etapa están dirigidos a los niveles directivos o mandos intermedios, que deben tomar decisiones de carácter más táctico o estratégico, en contraste con el carácter más técnico de la etapa anterior de analítica de datos, cuyos usuarios tendrán un perfil más de ingeniería o científico y utilizarán unos métodos de análisis más específicos.
- **Seguridad y Monitorización de Datos:** Esta etapa no es una etapa en sí, sino más bien un elemento transversal a todo sistema informático. Como tal, este gemelo operativo deberá tomar todas las medidas de seguridad necesarias para asegurar la protección adecuada de los activos de datos, controlar el acceso al sistema, proporcionar trazabilidad y auditoría y, especialmente en este caso por su impacto operativo, asegurar la disponibilidad del sistema. No obstante, no es objetivo de este trabajo abordar la cuestión de la seguridad de un sistema informático y sus comunicaciones, sea del ámbito de Defensa o no. En todo caso se asume que la seguridad, tanto del sistema de información como de las comunicaciones, son una premisa subyacente a la infraestructura sobre la que se desplegaría este sistema. De hecho, cuando en el capítulo anterior se trató el tema de los sistemas de comunicaciones navales, se pudo ver que la ciberseguridad es un requisito básico de diseño. Aun así, para abordar esta cuestión se debería hacer un análisis de riesgos del sistema, comenzando con la identificación de los activos, luego las amenazas que podrían materializarse teniendo en cuenta las diferentes dimensiones de la seguridad (disponibilidad, integridad y confidencialidad) para cada activo, y analizando la probabilidad y el impacto de que se materializaran, obteniendo así unos valores de riesgo. En base a estos, se determinarían las medidas necesarias a implementar para eliminarlos o mitigarlos.

En la *Figura 3-2* se muestra un diagrama de bloques que describe de forma sencilla la arquitectura y funcionalidad de esta parte del sistema acompañando logotipos de las instituciones que proporcionan los datos, formatos de datos y productos muy conocidos que se utilizan en cada una de las fases. Esta imagen no pretende en ningún caso dar una solución cerrada, sino simplemente dar ideas de por donde se mueven las tecnologías en este tipo de sistemas. Por ejemplo, se incluye *Hadoop* para el almacenamiento al considerarse uno de los sistemas de archivos distribuidos más extendidos en los ecosistemas de *Big Data*, apropiado para implementar *Data Lakes* de alto rendimiento y como sistema base para otras capacidades o herramientas como *MapReduce* para paralelizar el procesamiento de datos, o *Kafka* como gestor de flujos de datos en tiempo real. En la parte de analítica se incluyen múltiples alternativas, cada

una con sus particularidades que las hacen un poco más adecuadas para determinadas tareas y menos para otras, pero estos detalles se salen del alcance de este trabajo.

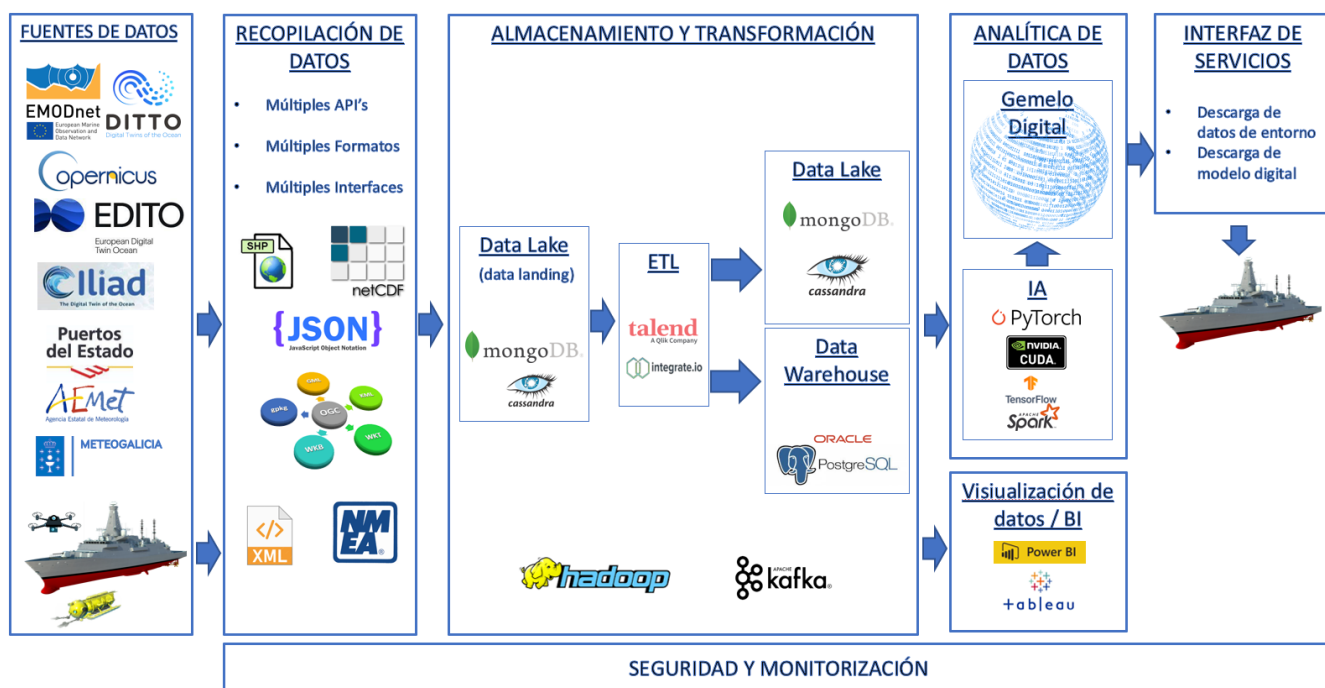


Figura 3-2: Arquitectura del sistema central (Nivel 1)

3.2.1 Información proporcionada por el gemelo digital

Una vez definida la arquitectura de este nivel, quedaría por concretar el tipo de información de entorno marítimo que se va a proporcionar y que puede contribuir a la mejora operativa de las plataformas. Hay que tener en cuenta que este Nivel 1 es en el que se fusiona toda la información de todas las fuentes disponibles, tanto institucionales como de plataformas propias, por lo que se constituirá como la referencia de todo el sistema. Por tanto, lo que se produzca en este nivel, y cómo se produzca, será el verdadero gemelo digital que proporcionará a las plataformas el modelo del entorno operativo marítimo y evolucionará conforme a los datos recibidos de las diferentes fuentes.

A continuación, se proporciona una lista de las variables que se han identificado como más adecuadas para la modelización del entorno marítimo, ya sea por su influencia en la operación, en las predicciones, por su disponibilidad o fiabilidad frente a otras variables:

- Datos y predicciones de condiciones meteorológicas en superficie
 - Dirección y velocidad del viento.
 - Presión atmosférica.
 - Temperatura.
 - Humedad relativa.
 - Fenómenos meteorológicos: lluvia, rayos, niebla, granizo, nieve, ...
 - Estado del cielo: despejado, altura de nubes, visibilidad.
- Datos de condiciones marítimas
 - Temperatura del agua.
 - Salinidad del agua (afecta mucho a la acústica y la detección de blancos).
 - Oleaje: frecuencia, altura, dirección.

- Estado de la marea.
- Corrientes.
- Datos de batimetría.
- Infraestructuras críticas.
- Datos de ayuda a la navegación (balizas, señales, faros, etc.).
- Localización de embarcaciones cercanas mediante datos AIS.

En este punto ya se tiene definido el qué, es decir, que información del entorno se maneja; lo siguiente es definir el cómo, esto es, la forma como se estructurará esta información para hacer análisis, transmitirla y representarla.

Para abordar esta cuestión se presentaron en el capítulo 2 los modelos de información geoespacial de rejillas jerárquicas, destacando entre ellos el modelo de celdas *H3* de *Uber*. Basándose en este modelo, el entorno se modelará como un conjunto de panales de celdillas hexagonales donde cada celda tendrá asignado un valor de la variable que representa. En la *Figura 3-3* se muestra un ejemplo de dos representaciones de diferentes variables propias de un modelado de entorno marítimo, en este caso salinidad y profundidad, con diferentes resoluciones espaciales.

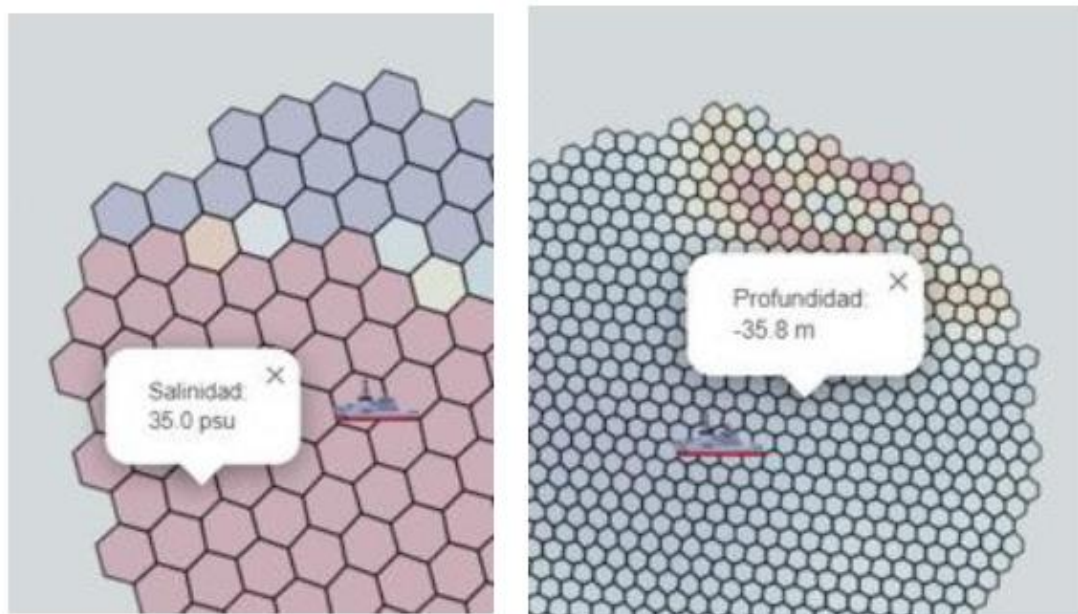


Figura 3-3: Ejemplo de representación de variables del entorno marítimo usando el modelo de celdas *H3* de *Uber* [61]

Se recalca la ventaja de estos sistemas jerárquicos ya que, aparte de permitir integrar datos de diferentes fuentes con diferentes resoluciones de una manera relativamente sencilla, permitirán controlar el volumen de información que se maneja, lo cual es relevante para las plataformas navales con conectividad limitada. Además, debido a las múltiples fuentes, formatos y naturalezas de la información, las transformaciones de cambios de resolución se anticipan como tareas recurrentes, pero en verdad esta es una de las capacidades que aportan potencia al modelo. De hecho, lo negativo será, más bien, que determinada información no requiera transformación alguna por no aportar una resolución muy fina, como por ejemplo algunos mapas satelitales.

Por otro lado, la principal desventaja del modelo *H3* para este sistema es que, si bien es muy apropiado para datos bidimensionales, no tienen un equivalente para el modelado de los datos en *3D*. Por tanto, para los casos de variables que varían de valor a diferentes alturas o profundidades, como

puede ser la presión o la temperatura, la visibilidad, las corrientes o la velocidad del viento, el modelo de celdas *H3* no lo contempla. No obstante, se pueden plantear formas de aprovechar las ventajas de los modelos de rejilla jerárquicos e implementar alguna aproximación que permita incorporar al sistema la dimensión de alturas y profundidades. Por ejemplo, las celdas *2D* de estos modelos jerárquicos se pueden interpretar también como pirámides hexagonales que se ensanchan o estrechan conforme se toma altura sobre la superficie o se desciende respectivamente. Teniendo esto en cuenta, los datos generados a más altura o más profundidad se asociarán a la celda sobre la que se proyectarían en el eje radial de la Tierra. La cuestión ahora radica en como incluir en el modelo la tercera dimensión de altura o profundidad. Para ello, se pueden elegir diversas alternativas, como que los propios datos codifiquen la altitud o profundidad en sus valores, o que se manejen conjuntos de rejillas *2D* de modo que cada rejilla represente un rango de altura o profundidad, lo que permitiría también trabajar con una mayor o menor resolución en esta dimensión según se requiera. Obviamente, esto no deja de ser una aproximación, pero teniendo en cuenta el tipo de variables que se manejan, así como la resolución y la naturaleza de los sistemas de medida que proporcionan los datos, esta información aproximada puede contribuir a la mejora de las estimaciones y de los modelos al aportar datos en esta tercera dimensión que de otra forma no estarían disponibles. No obstante, si al final se encontrara inviable o poco útil la incorporación de esta información *3D*, se podría trabajar solo con información *2D*.

Volviendo a los datos que el sistema central descargará de las plataformas, estos se van a detallar en los siguientes niveles. Lo que en este nivel se debe tener en cuenta es que, dependiendo del volumen de la información y la situación de las plataformas, algunos datos podrían actualizarse de forma frecuente, por ejemplo, a través de enlaces satelitales, y otros se descargarán cuando se tenga conexión a través de un enlace de banda ancha con las plataformas, por ejemplo, cuando estas se encuentran en puerto o cercanías.

3.2.2 Interfaces del componente

Para finalizar con la descripción del componente de Nivel 1 del sistema, se definen en este apartado las interfaces de entrada y salida de alto nivel, esto es, los tipos de información o peticiones que recibe sistema y los que proporciona a otros componentes. Además, se identifican las interfaces de comunicaciones a las que tendría acceso el sistema:

- Entradas:
 - Datos meteorológicos y marítimos obtenidos a través de servicios web institucionales y fuentes abiertas o propietarias.
 - Datos locales procedentes de las plataformas embarcadas.
 - Solicitudes de información de entorno operativo marítimo: estimaciones/predicciones para una localización geográfica y un periodo de tiempo indicados.
- Salidas:
 - Estimaciones y predicciones de entorno operativo marítimo en una localización geográfica y un intervalo de tiempo indicados.
 - Modelo predictivo entrenado utilizable por plataformas embarcadas tipo navío, buque o fragata en una localización geográfica y un periodo de tiempo determinados.
- Sistemas de comunicaciones disponibles:
 - Conexión de banda ancha: internet, redes privadas de la organización.
 - Redes móviles terrestres.
 - Redes satelitales.
 - Radio enlaces dedicados (*HF*, *VHF*, *UHF*, etc.).

3.3 Nivel 2: Plataforma naval con capacidades medias de procesamiento de datos.

La plataforma naval a la que se hace referencia corresponderá a un buque de tamaño medio o grande, tipo fragata por ejemplo, que disponga, por un lado, de una instrumentación muy completa que incluya, aparte de un radar y sónar más o menos sofisticados y los sensores más relacionados con la navegación como *GPS*, girocompás, etc., una serie de sensores de tipo medioambiental como los que se han descrito en el capítulo 2, capaces de proporcionar información meteorológica y marina como temperaturas del aire y del agua, fenómenos meteorológicos (precipitaciones, niebla, rayos, etc.), visibilidad, altura de nubes, salinidad del agua, corrientes, velocidad y dirección del viento, cámaras en diferentes espectros (visual e IR principalmente), etc. Por otro lado, esta embarcación deberá disponer de una capacidad de procesamiento y almacenamiento media que permita que se ejecuten los siguientes procesos:

- a. Almacenamiento y procesado de los datos capturados del entorno en tiempo real a partir de los sensores de los sistemas propios de la plataforma y de las plataformas auxiliares desplegadas desde la primera (que correspondería al Nivel 3 y que se trata más adelante).

Gracias a estos datos, este Nivel 2 podrá obtener estimaciones de las condiciones de navegación en el momento actual y el futuro cercano basadas en datos estimados y reales. La capacidad de almacenamiento y procesamiento necesarias va a depender de los tipos de sensores, los formatos de datos y cómo se quiera integrar o fusionar la información para mejorar la consciencia situacional. Básicamente, se busca transformar los datos en información para mejorar la operatividad durante el desarrollo de la misión, ya sea para su uso en el modelo digital predictivo que se carga en la plataforma desde el Nivel 1, o para su uso independiente en este nivel. Estas tareas de transformación pueden requerir el uso de una potencia de cómputo muy elevada, por lo que, al final, se deberá hacer un estudio de las posibilidades, beneficios y requerimientos de las mismas, ponderar todo ello, y decidir qué se implementa a este nivel, y qué se dejará para que sea procesado de manera diferida en el Nivel 1.

Como ya se ha comentado, los datos en crudo obtenidos de la propia plataforma y los sensores, que pueden suponer un volumen elevado, son datos valiosos que pueden aprovecharse en el entrenamiento del gemelo digital, por lo que a nivel de plataforma deberán ser almacenados debidamente, para posteriormente descargarlos en el sistema central (en el siguiente capítulo se hará hincapié en esta cuestión al describir las estrategias y casos de uso de integración de la información).

Por otro lado, determinados datos capturados pueden ser procesados y convertidos en algún tipo de información resumida cuyo volumen no sea muy elevado. Esta información podrá ser transmitida en condiciones de conectividad limitada al sistema central (por ejemplo, vía satélite o mediante radios tácticas), donde se podrá incorporar al gemelo digital completo para mejorar los datos y predicciones y ponerlos a disposición de otras plataformas a través del servicio de descarga de datos del entorno.

- b. Ejecución del modelo predictivo del entorno marítimo creado en el sistema central y descargado de este en la plataforma (Gemelo Digital Local).

Este modelo se alimentará en su funcionamiento, por un lado, de los datos del entorno y predicciones también descargadas desde el Nivel 1 y, por otro lado, de datos locales del entorno, más o menos procesados tal y como se indicó en el punto anterior, generados por los sensores del propio buque y los producidos por las plataformas auxiliares que se puedan desplegar desde el mismo.

En la *Figura 3-4* se muestra la arquitectura de componentes en este Nivel 2. Como se puede ver, hay un sistema de almacenamiento y transformación como en el Nivel 1, pero en este caso será a menor escala ya que únicamente se almacenarán los datos generados por los sensores, ya sean de la plataforma misma o de las auxiliares. Las principales razones de usar una arquitectura similar a la del Nivel 1 es, en

primer lugar, que es una solución flexible y escalable y, en segundo lugar, que así se mantienen comunalidad entre elementos, lo que facilita la administración, gestión, evolución y aprendizaje.

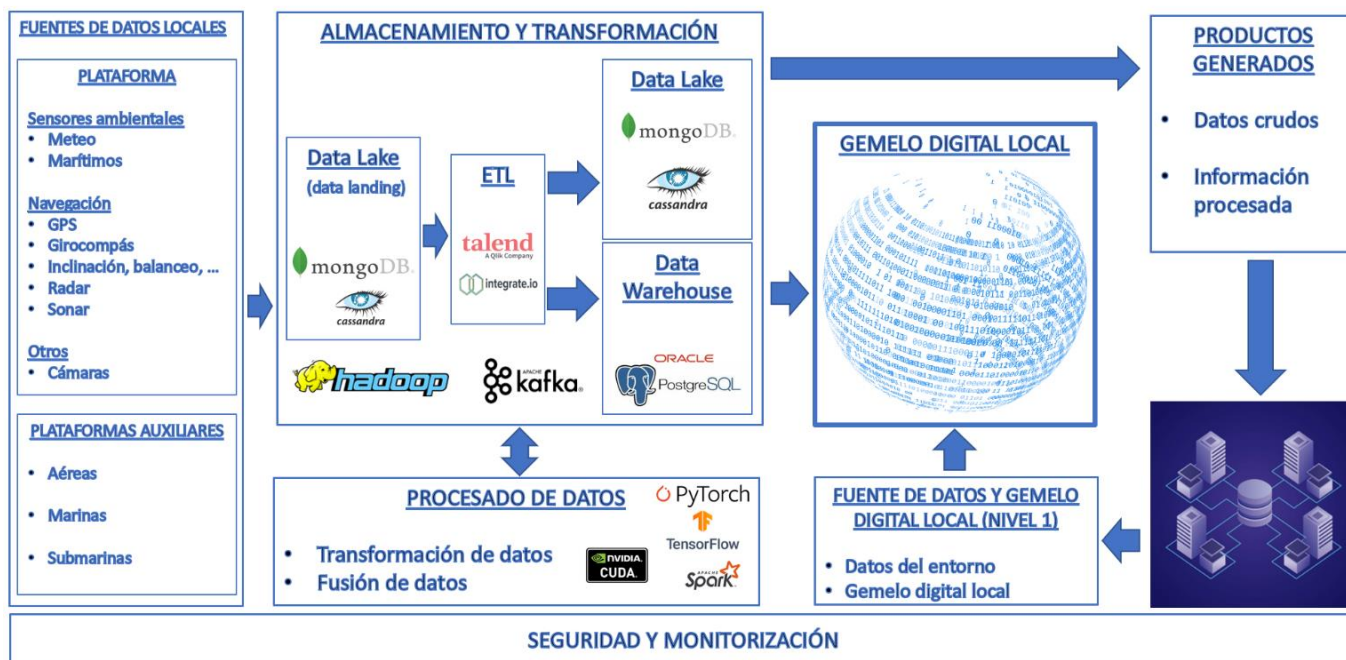


Figura 3-4: Arquitectura de integración de datos de la plataforma naval (Nivel 2)

El centro de este nivel, y así se ha querido reflejar en el diagrama, es el Gemelo Digital Local que, como se ha comentado, será alimentado a partir de los datos obtenidos de los sensores y de los datos provenientes del Nivel 1 (vía servicio), ya sea en una carga inicial o en una actualización posterior. También el Nivel 1 proporciona el modelo que representa el Gemelo Digital Local, que se ejecuta en la plataforma integrando los datos locales y proporcionando nuevas predicciones. El bloque de Procesado de Datos permite realizar transformaciones avanzadas de los datos o aplicar técnicas de fusión de datos que requieren cierta potencia de computación. Es aquí donde se haría la transformación de datos en información que se comentó anteriormente, que luego podría enviarse al Sistema Central para su integración y mejora a corto plazo de los datos proporcionados por el Gemelo Digital Central a las diferentes plataformas que soliciten actualizaciones de las condiciones de operación en un entorno indicado. En la parte de abajo del diagrama se vuelve a incluir el bloque transversal de Seguridad y Monitorización, dejando patente que todo el sistema debe prestar especial atención a estos aspectos, especialmente el primero de ellos.

3.3.1 Información proporcionada por el Nivel 2

Como ya se ha indicado más arriba, se genera dos clases de productos en el Nivel 2:

- Los datos en crudo que se descargarán en el Sistema Central cuando exista una comunicación de ancho de banda grande, y que será información que podrá estar en cualquier formato específico de los sensores.
- Los informes procesados elaborados a partir de los datos en crudo dentro de la propia plataforma. Esta información se transmitirá normalmente a través de enlaces tácticos y satélites, con un ancho de banda limitado, por lo que se requiere que sean compactos. Para ello se recurrirá a la solución ya indicada anteriormente de modelos de celdas *H3* de Uber, con varias capas en el caso de que se decida trabajado con información en *3D*.

El listado de variables que se consideran útiles para su envío, ya sea en crudo o procesados incluiría las mismas que se indicaron en el Nivel 1 (ver punto 3.1.1.1) que se refieren a condiciones meteorológicas y marítimas, y las siguientes propias de la plataforma:

- Datos cinemáticos de la plataforma:
 - Posición, velocidad y orientación del barco en los 3 ejes (rumbo, inclinación y balanceo).
- Fecha y hora (en realidad todos los datos tendrán que proporcionar referencia temporal).
- Datos de sonar.
 - Batimetría.
 - Objetos detectados submarinos.
- Datos de radar.
 - Objetos detectados en superficie/aire.
- Otros datos.
 - Imágenes/videos de sensores electroópticos (EO).
 - Objetos detectados vía sensores EO.

Puede llamar la atención que se hayan incluido los objetos detectados por sónar o radar. En principio podría pensarse que no tiene mucha relación con el modelado del entorno marítimo, pero podría ser una información útil a la hora de fusionar datos en el Sistema Central provenientes de diferentes fuentes. Además, puede tener un valor táctico interesante. No obstante, podría ser una información cuya transmisión fuera opcional y configurable en el sistema.

3.3.2 Interfaces del componente

Se termina con la descripción del componente de Nivel 2 definiendo en este apartado las interfaces de entrada y salida de alto nivel y las interfaces de comunicaciones a las que tendría acceso el sistema:

- Entradas:
 - Predicción de condiciones de navegación en una región geográfica y un intervalo de tiempo indicados procedentes del Nivel 1, bajo petición.
 - Modelo predictivo para uso local (Gemelo Digital Local) entrenado en el Nivel 1.
 - Datos procedentes de sensores de las plataformas auxiliares desplegadas (Nivel 3).
- Salidas:
 - Datos procesados de sensores o informes (requerirá ancho de banda pequeño).
 - Datos en crudo capturados por los sensores (requerirá ancho de banda grande).
 - Peticiones de descarga de datos: datos del entorno, modelo predictivo.
- Sistemas de comunicaciones disponibles:
 - Redes móviles terrestres (cerca de puerto).
 - Radio enlaces propios:
 - Radios tácticas (entre plataformas).
 - Wifi.
 - Otros (p.ej. con las plataformas desplegadas).
 - Redes satelitales.

3.4 Nivel 3: plataformas auxiliares desplegadas desde la plataforma principal

Este nivel comprendería las plataformas auxiliares desplegadas desde el Nivel 2, entre las que se pueden incluir diversos tipos en función del medio en el que operen, su tamaño, su forma de impulsión, el tipo de control, la misión para la que se hayan diseñado, etc. Así, por ejemplo, se podría hacer una primera distinción entre plataformas aéreas o marinas, y estas últimas podrían a su vez ser de superficie o submarinas; por su parte, las áreas podrían ser de ala fija o de hélices. Todas ellas podrían ser plataformas tripuladas o no tripuladas, y entre estas últimas podrían ser autónomas o controladas de forma remota. Entre las posibles misiones, podría haber misiones de reconocimiento, de ataque o defensa, de trabajo o manipulación, etc. Además, normalmente, cuanto más grandes sean las plataformas auxiliares, mayor será el número de sensores que lleven instalados por razones de espacio, energía, seguridad, polivalencia, etc., y mayor será la autonomía de las mismas.

De cara a la arquitectura del sistema, todas estas plataformas auxiliares de Nivel 3 son en realidad fuentes adicionales de datos para el Nivel 2, tal y como se reflejó al describir la arquitectura de ese nivel. Además, posiblemente, la mayoría de la información que podrían aportar coincidiría con alguno de los tipos descritos en el Nivel 2, ya que se asume que una plataforma naval de tamaño medio o grande lleva una instrumentación muy completa. El mayor impacto en la arquitectura de estos elementos consistirá en dotar al Nivel 2 de la infraestructura necesaria para capturar, almacenar y procesar los datos generados por estas plataformas desplegadas, pero eso ya se incluyó en la arquitectura del Nivel 2. Otro elemento significativo que implican estas plataformas auxiliares a nivel de arquitectura del sistema estaría relacionado con las interfaces de comunicación con las mismas, que puede hacerse a través de diferentes tecnologías: las plataformas más grandes dispondrán de radios tácticas e interfaces con redes tácticas como *Link-16* o *22*; por su parte, las plataformas más pequeñas pueden tener interfaces de radiofrecuencia más específicas o basadas en las últimas tecnologías móviles como *5G*. En cualquier caso, la comunicación entre los Niveles 2 y 3 no es un factor crítico del sistema ya que se supone que son enlaces locales y, en el peor de los casos, un fallo en estas comunicaciones significará, en lo referente al gemelo digital, no disponer de datos adicionales, pero la propia plataforma principal tendrá la información de sus sensores. Por tanto, estas plataformas auxiliares se tratarían como unos sensores adicionales de la plataforma principal, con la ventaja de que pueden llegar allí donde esta quizá no pueda llegar y proporcionar datos que pueden ampliar y mejorar su consciencia situacional. Además, como esta información llegará antes o después al sistema central, también servirá para la mejora de la información proporcionada por el gemelo digital central al resto de plataformas, ya sea a corto o a largo plazo, en función de si se transmite a través del servicio de informes del Nivel 2, o bien se descarga al Nivel 1 cuando la plataforma vuelva a puerto.

3.4.1 Información proporcionada por el Nivel 3

Sin ánimo de repetir, las plataformas de Nivel 3 proporcionarán en general datos similares a los de Nivel 2, si bien podrán extender la cobertura en tres dimensiones, siendo la dimensión de altitud o profundidad, según el caso, la que puede aportar datos no disponibles desde la plataforma principal. El modelado e integración de estos datos será el que implique la necesidad de incluir una tercera dimensión en el modelo de celdas *H3* de Uber para diferentes altitudes o profundidades, creando un modelo de múltiples capas, como ya se indicó anteriormente.

Dependiendo del tipo de plataformas desplegadas, los tipos de información que podrían aportar al sistema útil para la operativa del Gemelo Digital serían:

- Plataformas aéreas: Aviones, helicópteros, *UAV* (*Unmanned Aerial Vehicle*) o *RPAS* (*Remotely Piloted Aircraft System*).
 - Imágenes aéreas (visual e IR): objetos de superficie, estado de la mar, ...

- Datos de condiciones meteorológicas a diferentes altitudes y distancias dentro de su alcance.
- Contactos radar.
- Plataformas marinas de superficie: USV (*Unmanned Surface Vehicle*), RHIB (*rigid-hulled inflatable boat*).
 - Imágenes de superficie.
 - Condiciones de la mar y de superficie.
- Plataformas submarinas: mini-submarinos o UUV (*Unmanned Underwater Vehicle*).
 - Imágenes submarinas.
 - Condiciones de la mar.
 - Contactos sonar.

3.4.2 Interfaces del componente

En la sección anterior en la que se ha descrito la arquitectura global ya se ha tratado esta cuestión, así que, a modo de resumen, se enumeran las interfaces disponibles: radios tácticas (*VHF*, *UHF*, ...), redes tácticas (*Link-11*, 16, 22) y tecnologías de radiofrecuencia específicas o basadas en 5G.

4 OPERACIÓN DEL SISTEMA

Este capítulo abordará la parte más paráctica de la solución propuesta. En la primera sección se van a describir los diferentes contextos operativos, o casos de uso de operación del sistema de gemelo digital propuesto, para lo cual se incluirán una serie de diagramas que proporcionen una visión más gráfica e intuitiva del funcionamiento. En la segunda sección, se incluirá una breve discusión de los principales pros y contras de la solución propuesta por considerarse útil para ayudar en la toma de decisiones de cara a la futura implementación de un sistema de estas características.

4.1 Operación del gemelo digital: casos de uso de integración

El objetivo de esta sección es detallar los diferentes contextos en los que puede operar el sistema y cual será la forma de funcionar para, por un lado, aprovechar todas las posibilidades del modelo propuesto y, por otro lado, lidiar con las diferentes limitaciones existentes. Puesto que en el capítulo anterior ya se ha presentado la arquitectura de los diferentes niveles proporcionando diagramas de bloques que reflejan su operativa, y se ha descrito de forma detallada el funcionamiento de cada uno, se considera redundante presentar los casos de uso “internos” de cada nivel y se ha decidido abordar directamente los casos de integración entre ellos.

Inicialmente se barajó la posibilidad de describir los casos de uso mediante diagramas de interacción *UML (Unified Modelling Language)* con una secuenciación detallada, pero finalmente se consideró que, dado el nivel conceptual en el que se ha enmarcado el trabajo, no se dispone de unos requisitos lo suficientemente definidos y detallados para traducirlos en ese tipo de diagramas. Por este motivo, se ha optado por unos diagramas de casos de uso de nivel más funcional.

Aunque las comunicaciones en el entorno marítimo han avanzado considerablemente y su disponibilidad y capacidad es cada vez mayor, siguen existiendo limitaciones en estos aspectos que impiden que se pueda desplegar un gemelo digital del entorno marítimo que en tiempo real evolucione conforme a toda la información de las múltiples fuentes disponibles.

Las plataformas navales son cada día más complejas incluyendo multitud de sensores que proporcionan información de casi cualquier variable medible, entre ellas, las relacionadas con el entorno marítimo, y cuyo volumen aumenta aún más con las posibles plataformas auxiliares desplegables desde ellas. Es prácticamente imposible procesar todos estos datos en una plataforma marítima. Se requiere un potente *CPD* para utilizarlos en la evolución y desarrollo de nuevos modelos más precisos. Además, para mejorar los modelos es necesario integrar multitud de datos de multitud de plataformas, por lo que tampoco tiene mucho sentido el procesamiento masivo en la propia plataforma. Pero, por otro lado, tampoco pueden transmitirse todos estos datos, ya sea de forma *online* u *offline*, a los *CPDs* terrestres a través de los enlaces de comunicaciones navales satelitales o tácticos. Sin embargo, los datos son valiosos y no deberían perderse ya que podrán aprovecharse como datos históricos en el entrenamiento de nuevos modelos en los *CPDs*.

Teniendo en cuenta todo lo anterior, se debe plantear una estrategia de integración de los datos existentes en todo el sistema que contemple la conservación de la información y su explotación donde pueda llevarse a cabo y proporcionar valor añadido. La estrategia se basará principalmente en la disponibilidad y capacidad de comunicaciones entre el Nivel 1 y el Nivel 2, que dará lugar a tres escenarios, dentro de los cuales se identifican diferentes casos de uso:

- **Escenario 1: Sin restricciones de conectividad entre Nivel 1 y Nivel 2**
 - **Caso de uso 1.1:** Solicitud del Nivel 2 al Nivel 1 de modelos y datos de entorno.
 - **Caso de uso 1.2:** Descarga de datos crudos desde el Nivel 2 al Nivel 1.
- **Escenario 2: Conectividad restringida entre Nivel 1 y Nivel 2.**
 - **Caso de uso 2.1:** Solicitud del Nivel 2 al Nivel 1 de datos de entorno.
 - **Caso de uso 2.2:** Descarga de información filtrada/resumida desde el Nivel 2 al Nivel 1.
- **Escenario 3: Sin conectividad en la plataforma de Nivel 2 con el Nivel 1**
 - **Caso de uso 3.1:** La plataforma operará en modo *stand-alone* usando modelo pre-entrenado y datos propios.

4.1.1 Escenario 1: Sin restricciones de conectividad entre Nivel 1 y Nivel 2

Este escenario contempla la situación en la que la plataforma de Nivel 2 no tienen restricciones de conectividad con el sistema central terrestre. Esto generalmente ocurrirá cuando el buque esté en puerto o cerca del litoral y pueda utilizar servicios de comunicaciones propios o públicos: p.ej., si está en una base naval podría conectarse mediante cable, o si está dentro de alcance de la infraestructura terrestre, podría usar algún tipo de comunicaciones inalámbricas seguras de ancho de banda grande. Obviamente, deberá asegurarse la protección de las comunicaciones mediante las medidas que se consideren oportunas, como cifradores criptográficos.

En este escenario se va a asumir que la comunicación de ancho de banda grande es en los dos sentidos, aunque como se ha indicado previamente, la principal necesidad de envío masivo de datos va del Nivel 2 al Nivel 1, ya que es la plataforma desplegable la que deberá transmitir los datos en crudo registrados durante la misión al Nivel 1 para su uso en la evolución del gemelo digital. En cualquier caso, esto se describe en los siguientes apartados.

4.1.1.1 Caso de uso 1.1: Solicitud del Nivel 2 al Nivel 1 de modelos y datos de entorno operativo

En este caso de uso, que está dentro del escenario donde no hay restricciones de conectividad a comunicación, el sentido de la comunicación irá del Nivel 1 hacia Nivel 2, y tendrá dos propósitos principales, que serían como dos sub-casos de uso, aunque la operativa es muy parecida:

- i. **Actualización del modelo de predicción de la plataforma** a partir de un modelo más evolucionado en el sistema central. Como ya se indicó, es en este último donde se dispone de la potencia de cómputo necesaria para integrar todos los datos de fuentes externas y de todas las plataformas desplegables que se integren en el sistema y evolucionar o reentrenar los modelos. La operativa básica consistirá en que la plataforma iniciará un proceso de actualización del modelo y establecerá una conexión con el sistema central; se enviará la petición al sistema central indicando lo que se requiere, que puede ser un modelo para un uso en una zona determinada, un periodo determinado, o un modelo general; el sistema central, atenderá la petición y proporcionará el modelo, para lo cual puede ser necesario realizar una serie de configuraciones, ajustes, cálculos, etc. Este caso de uso aparece diagramado en la *Figura 4-1*.

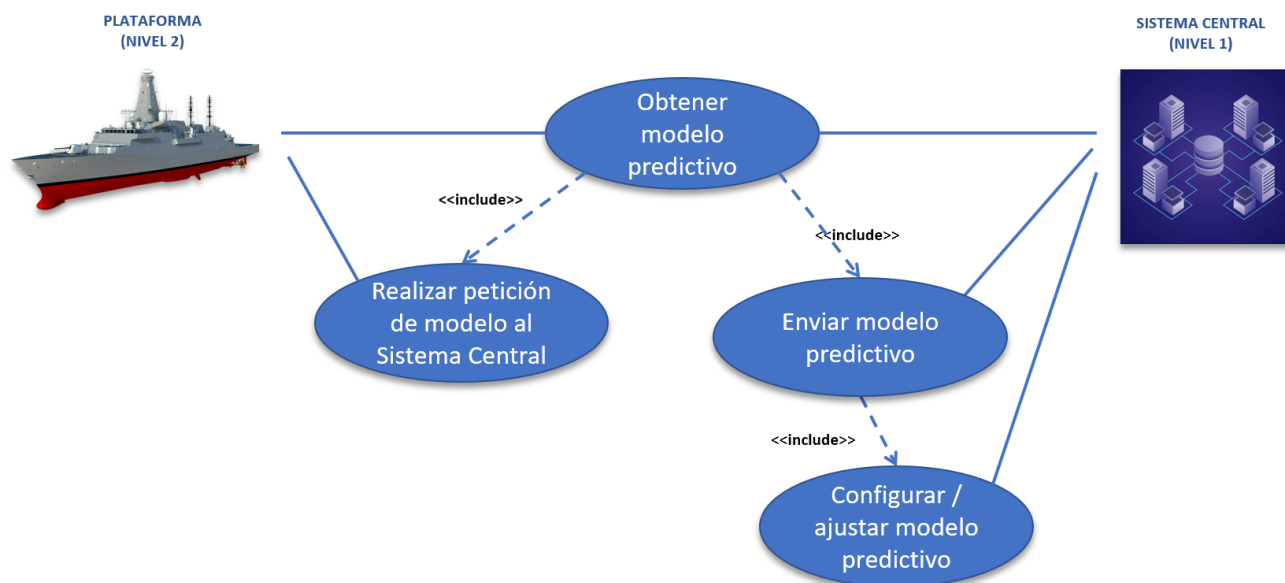


Figura 4-1: Caso de uso de actualización del modelo de gemelo digital para su ejecución en la plataforma (Escenario 1)

- ii. **Descarga de datos de un entorno operativo** donde se vaya a desarrollar una misión, lo que se podría denominar “datos de misión” para simplificar. Esta funcionalidad corresponde al servicio descrito en el apartado sobre la arquitectura global, mediante el cual una plataforma solicita al sistema central que le proporcione los datos del entorno en un área y durante un periodo de tiempo indicados para su uso operativo. No obstante, en este caso, puesto que se dispone de buena conectividad y se supone que se solicitaría previamente a una misión, podría ser una descarga más amplia, para un área o un periodo de tiempo mucho mayores, para que, en caso de que la plataforma se quede sin conectividad con el sistema central durante la misión, tenga un modelo del entorno en el que apoyarse, aunque no esté actualizado con los últimos datos recibidos e incorporados en el sistema central. En resumen, la idea es que la plataforma salga de puerto hacia una zona de operaciones con datos de las condiciones marítimas y climatológicas que se va a encontrar allí. La operativa por tanto sería: la plataforma establece una conexión con el servidor central para iniciar el proceso de carga de datos de misión; una vez conectado enviará la solicitud indicando el área y periodo de tiempo; el sistema central recibe la solicitud, y antes de proporcionar el modelo hará los cálculos necesarios para obtener las estimaciones/predicciones que correspondan. El diagrama de este caso de uso se presenta en la siguiente figura:

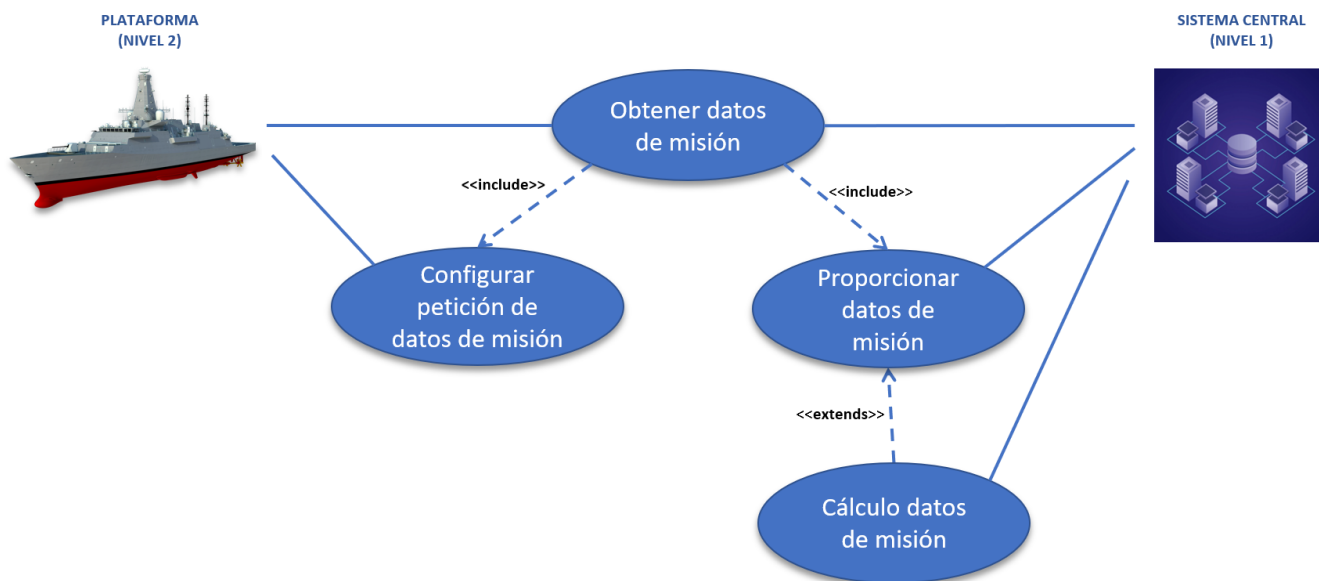


Figura 4-2: Caso de uso de descarga de datos del entorno operativo de la misión a la plataforma (Escenario 1)

4.1.1.2 Caso de uso 1.2: Descarga de datos crudos desde el Nivel 2 al Nivel 1

Este segundo caso de uso es el que corresponde a la descarga de los datos en crudo registrados durante la operación de la plataforma (Nivel 2) al sistema central (Nivel 1) para que este los almacene en su *Data Lake* para su posterior explotación y, principalmente, para la mejora de los modelos del gemelo digital, aunque también podrían aprovecharse estos datos para otros propósitos más tácticos o estratégicos.

Al describir más arriba la arquitectura del sistema, y más concretamente la del Nivel 2, ya se indicó que la plataforma registrará durante la misión multitud de datos, algunos de los cuales procesará con propósitos más operativos, y otros simplemente los almacenará para una explotación posterior en un CPD. Entre estos datos podrían encontrarse imágenes de alta resolución en diferentes espectros, videos, datos radar o sonar complejos, etc. Por tanto, el volumen y variabilidad de los datos puede ser bastante elevada, razón por la cual se requerirá una conexión de banda ancha para su transmisión, lo que, en principio, solo estará disponible en las cercanías del litoral o en los puertos o bases navales. Obviamente, este volumen de datos también podría exportarse a unidades extraíbles desde las que luego importar los datos en el sistema central, y podría haberse tratado como un caso de uso particular, pero tampoco se considera necesario ya que, en definitiva, lo que refleja el caso de uso es que no se está en un escenario de comunicaciones limitadas ni de incomunicación como los que se tratan en los siguientes apartados.

La operativa de la descarga consistirá en configurar primeramente la descarga, incluyendo los elementos que se consideren necesarios; posteriormente se establecerá una conexión con el servidor central, para lo que se implementarán todas las medidas de seguridad oportunas; el sistema central aceptará la descarga y recibirá los datos en el “*Landing Data Lake*”, para posteriormente procesarlos y explotarlos como corresponda.

El siguiente diagrama describe el caso de uso indicado:

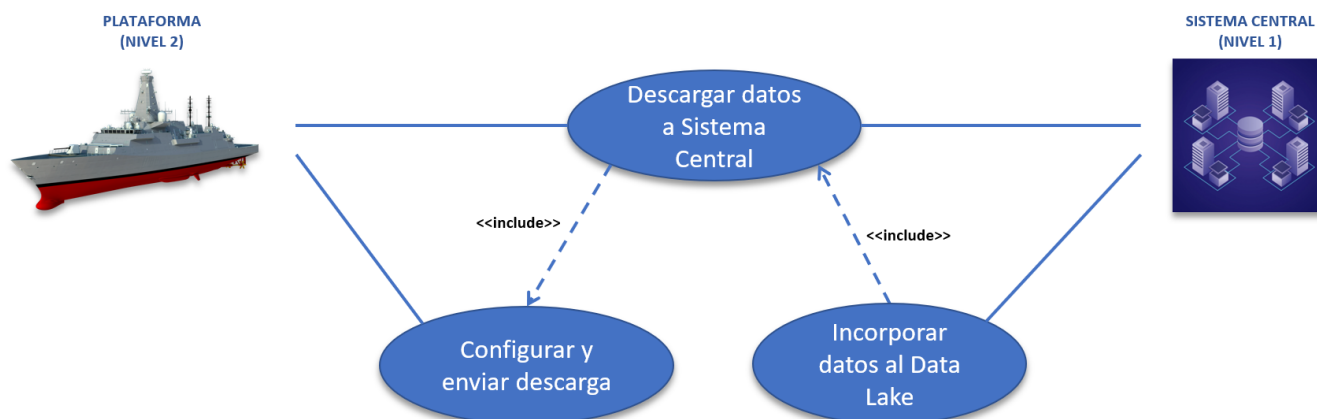


Figura 4-3: Caso de uso de descarga de datos del entorno operativo al sistema central (Escenario 1)

4.1.2 Escenario 2: Conectividad restringida entre Nivel 1 y Nivel 2.

En este escenario el buque se encontrará normalmente desplegado en un área de operaciones donde no habrá conectividad de banda ancha con el Sistema Central, pero sí habrá otras formas de conexión; en función de la lejanía y las capacidades de los sistemas de comunicaciones de la nave, se podrán establecer enlaces vía satélite, que es lo más común, o enlaces radio HF, que, como se ha visto en el capítulo anterior, cada vez proporcionan mejores características para la transmisión de datos, a más distancia y con mayor ancho de banda. En cualquier caso, el ancho de banda disponible es muy valioso y las velocidades de transmisión serán relativamente bajas, por lo que se hace necesario priorizar y limitar el tipo de información que se debe transmitir.

Teniendo en cuenta lo anterior, en la definición de las arquitecturas y funcionalidades del sistema se identificaron dos tipos de servicios de transmisión de información entre Nivel 2 y Nivel 1 para escenario de comunicaciones limitadas. Estos servicios eran:

- i. Servicio de solicitud de “datos de misión” (del entorno) para alimentar el Gemelo Digital Local con los datos más actualizados proporcionados por el Gemelo Digital Central.
- ii. Envío de información filtrada/resumida desde el buque al Sistema Central para su integración en el Gemelo Digital Central y mejorar los datos proporcionados por este a todos los usuarios basados en datos más recientes y precisos adquiridos por las diferentes plataformas desplegadas.

Estos servicios dan lugar a 2 casos de uso de integración de datos que se describen en los dos subapartados siguientes.

4.1.2.1 Caso de uso 2.1: Solicitud del Nivel 2 al Nivel 1 de descarga de datos de misión

Este caso de uso refleja la situación en la que la plataforma desplegada en un área de operaciones requiere una actualización de la estimación/predicción de las condiciones de navegación para mejorar su consciencia situacional y poder tomar las decisiones más adecuadas basándose en datos más recientes y precisos. Para ello, hará una solicitud al sistema central que se transmitirá a través del enlace vía satélite, o un enlace radio HF de datos (si dispone de esta capacidad). En esta solicitud, la plataforma podrá configurar determinados parámetros, como pueden ser el área para el cual requiere los datos, y el periodo de predicción máximo. La decisión sobre estos parámetros se basará en diferentes variables como, por ejemplo, la previsión de disponibilidad de comunicaciones, ya que, si se prevé que pueda

haber una interrupción de las mismas por un tiempo prolongado, será conveniente precargar el modelo con la mayor cantidad de información para disponer de datos en los que basar la evolución del modelo predictivo de la plataforma durante ese periodo. Otra variable a tener en cuenta será la capacidad del enlace a través del cual se solicita la información, ya que, si la velocidad de transmisión es muy baja, puede no ser viable solicitar información muy extensa y para un periodo de tiempo muy dilatado. Como se puede ver, la operativa del sistema está condicionada por las capacidades de comunicaciones, las cuales se encuentran a su vez condicionadas por la situación táctica o estratégica del escenario en el que se desarrolla una operación.

El diagrama de caso de uso de esta operativa es exactamente igual al que se representa en el Escenario I para la carga de datos de misión. Lo que variaría en este caso es el enlace a través del cual se desarrolla la petición. No obstante, se incluye el diagrama aquí otra vez por motivos de claridad:

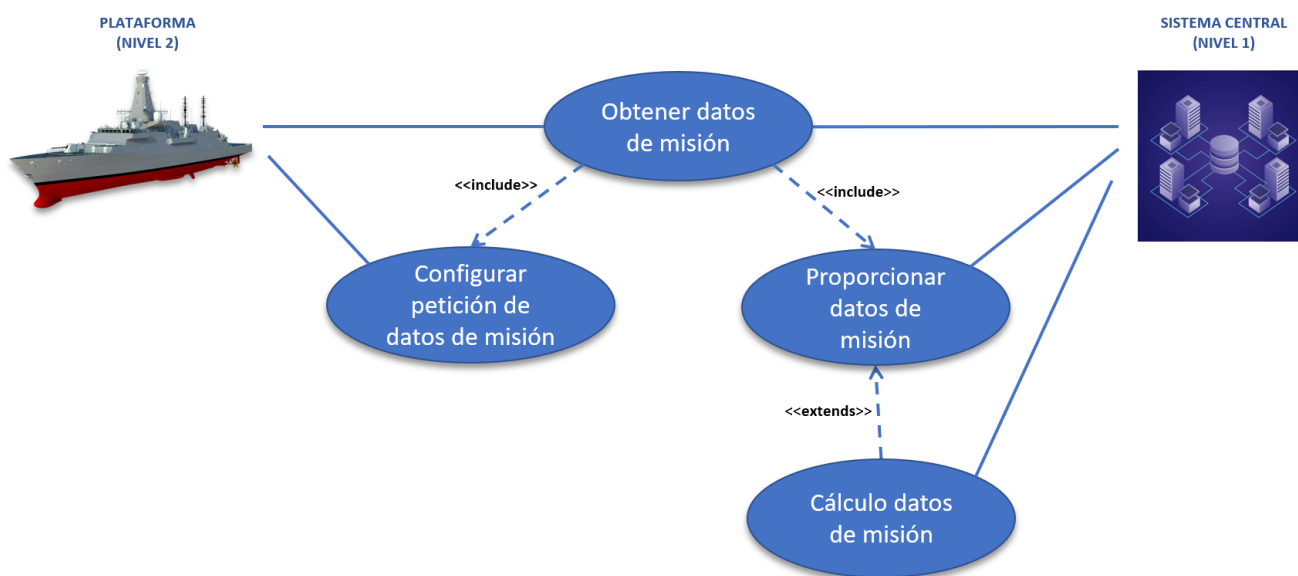


Figura 4-4: Caso de uso de descarga de datos del entorno operativo de la misión a la plataforma (Escenario 2).

4.1.2.2 Caso de uso 2.2: Descarga de información filtrada/resumida desde el Nivel 2 al Nivel 1

En este caso de uso, el objetivo es alimentar al Gemelo Digital Central con datos recientes obtenidos por las plataformas. Como ya se ha comentado, una plataforma puede generar una cantidad inmensa de datos. Sin embargo, en un escenario de conectividad limitada no es posible transmitir todos esos datos lo que, por otro lado, tampoco tendría mucho sentido ya que no sería muy factible procesarlos en tiempo real para incorporarlos al Gemelo Digital y mejorar de forma inmediata las predicciones. La idea es seleccionar determinados datos que han sido procesados en la plataforma y convertidos en información valiosa y resumida que sí pueda ser utilizada en el Gemelo Digital Central para mejorar las predicciones en el corto plazo y que estas puedan ser proporcionadas a todas las plataformas que soliciten datos actualizados. Esta información resumida se estará generando de forma continua y será poco viable transmitirla conforme esté disponible, ya que seguramente saturaría el enlace de comunicaciones. La idea será transmitirla de forma periódica, por ejemplo, cada x minutos, o cada hora, o cada x horas, enviando únicamente la información actualizada al instante de envío. También hay que tener en cuenta que la información, aunque resumida, puede ser muy extensa y puede ser conveniente enviar diferentes tipos de información con diferentes periodicidades. Como se puede ver, la estrategia de envío de información (qué y cada cuánto) va a venir condicionada, una vez más, por la capacidad disponible de las comunicaciones.

Se muestra a continuación el diagrama del caso de uso descrito:

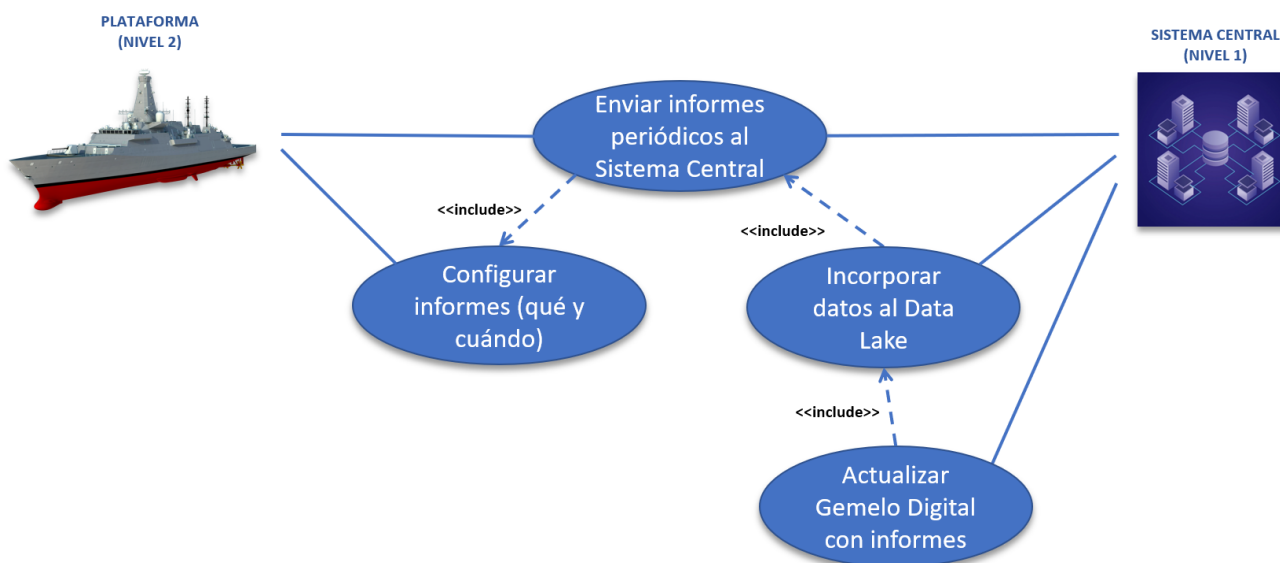


Figura 4-5: Caso de uso de envío de informes de las plataformas desplegadas al sistema central (Escenario 2).

4.1.3 Escenario 3: Sin conectividad en la plataforma de Nivel 2 con el Nivel 1

En este tercer escenario, el Nivel 1 no va a estar accesible por la plataforma, lo que puede deberse a diferentes causas: por ejemplo, puede estar inoperativo por avería el sistema de comunicaciones vía satélite del buque; o puede estar caído el Sistema Central; o puede que exista perturbación electromagnética (*jamming*) en el área de operación que impida que las comunicaciones funcionen correctamente, etc.

En este contexto de no disponibilidad de comunicaciones con el Sistema Central, obviamente, no habrá ningún tipo de integración de datos entre el Nivel 1 y el Nivel 2, y la plataforma deberá funcionar en un modo *stand-alone*. La integración de datos será cuestión única de la plataforma principal a partir de sus propios sensores y de los datos generados por las plataformas auxiliares desplegadas, si es que estas últimas pueden operar y sus datos ser transmitidos a la plataforma, ya que también podrían verse afectadas por problemas en las comunicaciones locales. La idea es que todos los datos generados disponibles se registren en el sistema de almacenamiento de la plataforma para procesarlos en el momento, o proporcionárselos posteriormente al Sistema Central para su explotación. Dentro de la plataforma, los datos se procesarán de acuerdo a las posibilidades disponibles en cuanto a capacidad de cómputo, y se utilizarán para alimentar el Gemelo Digital Local, de modo que este pueda integrar estos datos y producir predicciones más precisas.

En el siguiente diagrama se representa este caso de uso:

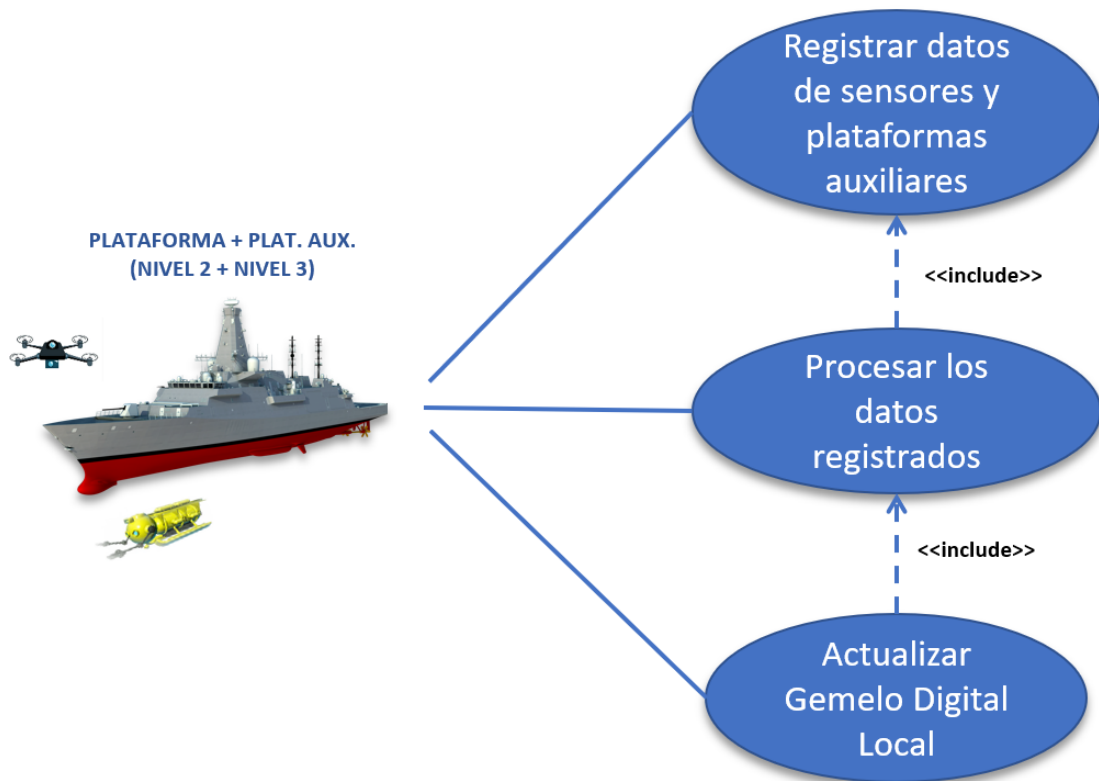


Figura 4-6: Caso de uso de funcionamiento de la plataforma en modo stand-alone (Escenario 3)

4.2 Discusión de la solución

Esta sección pretende hacer un pequeño análisis de pros y contras asociados a la solución propuesta. Puesto que se está hablando de un modelo conceptual, es complicado ir más allá de la identificación de aquellos aspectos que aportan valor y aquellos que pueden dificultar o suponer un riesgo para el desarrollo y operación de esta solución.

Como aspectos positivos de esta solución se destacan los siguientes:

- Permite disponer de un modelo del entorno en las plataformas desplegadas (gemelo digital local) que proporciona estimaciones de las condiciones de navegación aun sin tener conectividad con el sistema central, siempre y cuando se haya descargado de este último en algún momento previo.
- Capacidad de actualizar el gemelo digital local cuando se disponga de conectividad.
- Capacidad de actualizar las estimaciones de las condiciones de navegación en cualquier momento en el que haya conectividad.
- Capacidad de realimentar al gemelo digital central información real actualizada para la mejora de las predicciones y su actualización a otras plataformas desplegadas.
- Capacidad de controlar el volumen de información que se intercambia en escenarios de conectividad limitada, pudiendo seleccionar área y plazo de predicción, tipo de información, resolución, frecuencia de actualizaciones, etc.
- Capacidad de alimentar el gemelo digital central con gran cantidad de datos en crudo reales, obtenidos por las plataformas durante su operación, útiles para la mejora de los modelos o el desarrollo de otros nuevos.

Entre las principales limitaciones o dificultades se identifican las siguientes:

- Dependencia de un sistema central para la obtención de modelos y datos de los entornos de operación actualizados. Si el sistema central deja de funcionar, las plataformas desplegadas no podrán actualizar los modelos y escenarios y deberán trabajar con los últimos modelos que hubieran descargado del sistema central, actualizados sólo con los datos locales que producen, puesto que no se contempla inicialmente la conexión directa con las fuentes de datos institucionales. Dependiendo del tiempo de indisponibilidad del sistema central, esto podría influir negativamente en la operatividad de las plataformas.
- Altos requisitos de capacidad informática instalada en las plataformas. Si se quiere aprovechar al máximo la información generada en las plataformas para alimentar el gemelo digital local y el central, las plataformas deberán disponer de capacidad suficiente para el almacenamiento y procesamiento de toda la información generada por los sensores de la propia plataforma y las auxiliares.
- Muy relacionado con el punto anterior, para poder aportar información valiosa a los modelos, las plataformas desplegadas deben incorporar una instrumentación muy potente que permita obtener datos fiables de todas las variables que puedan llegar a ser relevantes para la operación y el desarrollo de los modelos.
- Complejidad de los modelos predictivos. Las diferentes fuentes de información que pueden ser tanto externas como locales, de naturalezas diferentes, con múltiples tipos y formatos de datos, distintas frecuencias de actualización, procedentes de múltiples áreas geográficas, con diferentes resoluciones, etc., suponen un desafío muy complejo para el desarrollo de modelos basados en IA. Se deberán desarrollar diferentes modelos para diferentes procesos, propósitos y plataformas. Por ejemplo, probablemente no será igual el gemelo digital que corre en el sistema central que los que corren en las plataformas. El entrenamiento de todos estos modelos requerirá un trabajo enorme, con muchos ciclos de prueba y error, y unos requisitos computacionales enormes.

Como se puede ver, la solución propuesta es tan ambiciosa y prometedora como compleja y exigente. Por ello, una futura implementación basada en esta solución conceptual debería abordarse mediante una aproximación de tipo incremental, analizando y priorizando los diferentes requisitos y funcionalidades, su complejidad y viabilidad, siempre con la vista puesta en la mejora de la operatividad de las plataformas que despliegan en las misiones.

5 CONCLUSIONES Y LÍNEAS FUTURAS

En este trabajo se ha abordado, desde un punto de vista conceptual o funcional, la problemática asociada a la implementación de un Gemelo Digital Marítimo en un entorno operativo donde desplegarían plataformas navales para llevar a cabo sus misiones y en el que existen limitaciones de comunicaciones que dificultan mantener actualizado el Gemelo Digital.

Se ha podido ver que los denominados Gemelos Digitales implican una complejidad muy elevada, no solo en lo referente a las tecnologías necesarias para su implementación, sino también respecto a las condiciones operativas en las que se pueden desplegar, y donde el medio marino es especialmente complicado.

Abordar un proyecto de estas características va a requerir considerar una gran multitud de aspectos, algunos de los cuales pueden ser críticos para el éxito de la solución, y otros muchos que obligarán a tomar decisiones que lleven a soluciones de compromiso. No obstante, las tecnologías e infraestructuras, tanto de sistemas de información como de comunicaciones, están avanzando rápidamente y es posible que algunas limitaciones existentes actualmente dejen de serlo en los próximos años.

Por otro lado, queda patente que los datos masivos son los cimientos sobre los que se apoya la tecnología del Gemelo Digital. En este sentido y, afortunadamente, en la actualidad se está trabajando de forma importante en varios proyectos de ámbito internacional que abordan el desarrollo de gemelos digitales oceánicos que pondrán a disposición de quien lo necesite datos, predicciones y modelos cada vez más completos y precisos. No obstante, parece que aún falta para llegar a este punto y se anticipa que va a ser en los próximos años cuando se produzca un salto cuantitativo y cualitativo.

Teniendo en cuenta lo anterior, se ha planteado una solución conceptual basada en una arquitectura de diferentes niveles acompañada de unas estrategias de operación que permiten aprovechar al máximo los datos disponibles en los diferentes escenarios operativos con las limitaciones indicadas, sobre todo en lo referente a las comunicaciones. Una implementación basada en esta solución no sólo mejoraría la consciencia situacional a las plataformas en esos escenarios, sino que además permitiría evolucionar el gemelo digital para mejorar sus predicciones a medida que se vayan incorporando más datos generados por las plataformas en el transcurso de sus misiones. En este sentido, se considera que se han cumplido los objetivos de este trabajo y que la solución propuesta puede usarse como marco de referencia para una futura implementación.

Como líneas de trabajo a seguir a partir de aquí, se plantean numerosas posibilidades debido a la amplitud de la materia, escenarios y tecnologías implicadas. No obstante, dada a la complejidad y el difícil acceso al entorno real, sería conveniente enfocar los trabajos principalmente en herramientas de simulación. Por ejemplo, una línea de trabajo podría enfocarse en el análisis del modelo considerando los aspectos más cuantitativos de la integración, analizando las necesidades y limitaciones de volumen de datos a transmitir, que dependerá de la resolución de los modelos con los que se quiera trabajar y de la información que se considera necesaria o relevante, así como el ancho de banda proporcionado por los diferentes sistemas de comunicaciones disponibles.

Otras líneas de trabajo podrían abordar cuestiones más relacionadas con los modelos predictivos en este contexto, donde también se abre un enorme campo de posibilidades. Por ejemplo, se podrían estudiar los diferentes casos de integración de datos locales que conlleva la solución. Se podría estudiar el grado de mejora de las predicciones del gemelo digital en el Nivel 1 producido por la integración de la información proveniente de informes resumidos de las plataformas. También se podría estudiar el funcionamiento del modelo entrenado en el Nivel 1 que se ejecuta en la plataforma, alimentado inicialmente por datos del Nivel 1, y luego actualizado únicamente con los datos obtenidos localmente.

En cuestiones relacionadas con la forma de representar el entorno y los formatos de datos, podrían estudiarse diferentes alternativas para la incorporación y explotación de datos *3D* en el sistema que el modelo de celdas *H3* de Uber no soporta. Podría incluso estudiarse si merece la pena incorporar esta información *3D* o no va a aportar mejoras a las predicciones del modelo o la consciencia situacional de las plataformas.

También podrían estudiarse las diferentes soluciones o ecosistemas *Big Data* para el almacenamiento o la analítica de datos. O podría estudiarse la integración de datos de múltiples plataformas desplegadas en la misma área sin pasar por el sistema central. En definitiva, existen multitud de posibilidades donde trabajar para llegar a desarrollar una solución real operativa de gemelo digital del entorno marítimo que permita a las plataformas llevar a cabo sus misiones de forma más segura, eficaz y eficiente.

Para terminar, y a título personal, considero que el desarrollo de este trabajo me ha permitido acercarme a un campo de conocimiento, el gemelo digital, que creo que va a marcar los próximos años y en el que confluyen multitud de disciplinas tecnológicas, que además son llevadas al extremo, lo que crea un caldo de cultivo muy adecuado para su investigación y desarrollo. También me ha posibilitado conocer un poco más la problemática asociada al mundo naval, en el que no contaba con experiencia, pues he trabajado, fundamentalmente, en el mundo aeronáutico. En este sentido, el desarrollo de este trabajo me ha sugerido multitud de ideas que podrían aprovecharse para el entorno aeronáutico, donde la climatología condiciona fuertemente el desarrollo de las misiones y disponer de un “gemelo de condiciones de vuelo” podría ser muy interesante. El principal problema es que, en las plataformas aéreas actuales, difícilmente se puede instalar una potencia de cómputo similar a la que se puede instalar en una fragata, aunque quizás esto sea una cuestión de tiempo, o se puedan desarrollar soluciones más limitadas, pero también válidas.

BIBLIOGRAFÍA

- [1] «La Transformación Digital de la Industria Española. Informe Preliminar». Accedido: 8 de enero de 2024. [En línea]. <https://www.industriaconectada40.gob.es/SiteCollectionDocuments/informe-industria-conectada40.pdf>
- [2] R. Blanco, «LA INDUSTRIA 4.0: EL ESTADO DE LA CUESTIÓN». Ministerio de Industria, Tecnología y Turismo. Accedido: 8 de enero de 2024. [En línea]. <https://www.mintur.gob.es/Publicaciones/Publicacionesperiodicas/EconomiaIndustrial/RevistaEconomiaIndustrial/406/BLANCO,%20FONTRODONA%20Y%20POVEDA.pdf>
- [3] «Copernicus Marine Data Store». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://data.marine.copernicus.eu/products>
- [4] «¿Qué es un gemelo digital? | IBM». Accedido: 8 de enero de 2024. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/what-is-a-digital-twin>
- [5] Á. G. de Ágreda y I. M. Herranz, «Usos militares de la inteligencia artificial, la automatización y la robótica (IAA&R)».
- [6] A. Rasheed, O. San, y T. Kvamsdal, «Digital Twin: Values, Challenges and Enablers». arXiv, 3 de octubre de 2019. doi: 10.48550/arXiv.1910.01719.
- [7] «What is data assimilation?», DARC. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://research.reading.ac.uk/met-darc/aboutus/what-is-data-assimilation/>
- [8] M. M. Salim, «Digital twin technology: An evaluation». 2023.
- [9] Sahilbhutada, «Digital twin technology and machine learning», Medium. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://medium.com/@sahilbhutada28/digital-twin-technology-and-machine-learning-f68c31ee6ef>
- [10] C. Verdouw, B. Tekinerdogan, A. Beulens, y S. Wolfert, «Digital twins in smart farming», *Agricultural Systems*, vol. 189, p. 103046, abr. 2021, doi: 10.1016/j.agsy.2020.103046.
- [11] «European Digital Twin Ocean - Powered by EDITO», EDITO-Infra. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://edito-infra.eu/>
- [12] «Destino Tierra | Configurar el futuro digital de Europa». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://digital-strategy.ec.europa.eu/es/policies/destination-earth>
- [13] «Digital Twins of The Ocean - The Iliad Project». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://ocean-twin.eu/>
- [14] «Who we are. The Ocean Decade in a nutshell», Ocean Decade. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://oceandecade.org/who-we-are/>
- [15] «Ocean Decade Action Factsheet: Digital Twins of the Ocean (DITTO)». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://oceanexpert.org/document/29741>
- [16] «Open Geospatial Consortium - Home Page», Open Geospatial Consortium. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.ogc.org/>
- [17] A. Morales, «Los 10 formatos GIS vectoriales más populares», MappingGIS. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://mappinggis.com/2013/11/los-formatos-gis-vectoriales-mas-populares/>
- [18] «Ejemplos JSON archivo datos. Cambio formato convertidor on line XML a JSON y al revés. Minificar (CU01214F)». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: https://www.aprenderaprogramar.com/index.php?option=com_content&view=article&id=957:ejemplos-json-archivo-datos-cambio-formato-convertidor-on-line-xml-a-json-y-al-reves-minificar-cu01214f&catid=83&Itemid=212
- [19] «OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard», Open Geospatial Consortium. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.ogc.org/standard/eo-geojson/>
- [20] G. Poole, «TopoJSON, GeoJSON, and Projections: Developing Interactive Map Visuals for Web Apps», Medium. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en:

- <https://betterprogramming.pub/topojson-geojson-and-projections-developing-interactive-map-visuals-for-web-apps-79e2d50abb19>
- [21] «Formats | EUMETSAT». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.eumetsat.int/formats>
- [22] «Shapefile», *Wikipedia, la enciclopedia libre*. 10 de agosto de 2023. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://es.wikipedia.org/w/index.php?title=Shapefile&oldid=152978164>
- [23] A. Morales, «GeoPackage para novatos: ventajas y uso en ArcGIS, QGIS y GeoServer», MappingGIS. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://mappinggis.com/2017/06/geopackage-para-novatos-uso-en-arcgis-qgis-publicacion-en-geoserver/>
- [24] «IOC Ocean Data and Information System Catalogue». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://catalogue.odis.org/>
- [25] «IOC Oceanographic Data Exchange Policy and Terms of Use (2023)». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: https://www.iode.org/index.php?option=com_content&view=article&id=51:ioc-oceanographic-data-exchange-policy&catid=24&Itemid=100040
- [26] «European Marine Observation and Data Network (EMODnet)». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://emodnet.ec.europa.eu/en>
- [27] «EMODnet web service documentation», European Marine Observation and Data Network (EMODnet). Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://emodnet.ec.europa.eu/en/emodnet-web-service-documentation>
- [28] C. D. S. Ecosystem, «Explore data | Copernicus Data Space Ecosystem». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://dataspace.copernicus.eu/explore-data>
- [29] «AEMET OpenData». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://opendata.aemet.es/centrodedescargas/productosAEMET?>
- [30] Ministerio de Agricultura, Alimentación y Medio Ambiente, *Resolución de 30 de diciembre de 2015, de la Agencia Estatal de Meteorología, por la que se establecen los precios públicos que han de regir la prestación de servicios meteorológicos y climatológicos*, vol. BOE-A-2016-111. 2016, pp. 649-662. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.boe.es/eli/es/res/2015/12/30/1>
- [31] «Predicción de oleaje, nivel del mar; Boyas y mareógrafos | puertos.es». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.puertos.es/es-es/oceanografia/Paginas/portus.aspx>
- [32] «SERVICIO ONLINE DE DESCARGA DE DATOS OCEANO-METEOROLÓGICOS DE PUERTOS DEL ESTADO.». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: https://bancodatos.puertos.es/BD/peticiones/Manual_Uusuario_Descargaportus.pdf
- [33] «Introducción al uso de los servicios Thredds. MeteoGalicia.» Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: https://www.meteogalicia.gal/datosred/infoweb/numerico/thredds/Manual_uso_Thredds.pdf
- [34] «Servidor THREDDS de MeteoGalicia» Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: https://www.meteogalicia.gal/modelos/thredds/threddsIndex.action?request_locale=es
- [35] «Campbell - Rugged Monitoring - Measurement and control instrumentation for any application». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <http://www.campbellsci.com/>
- [36] «Vaisala - a global leader in environmental and industrial measurement |». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.vaisala.com/en>
- [37] «Observator.com - Home», Observator. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://observator.com/>
- [38] «Todo sobre las redes NMEA 0183 y NMEA 2000 - Digital Yacht», Digital Yacht España. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://digitalyacht.es/nmea0183->

- [y-nmea2000/](#)
- [39] «CURRENT INDICATOR CI-68 | Current Indicator | Products | FURUNO». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.furuno.com/en/products/currentIndicator/CI-68>
 - [40] «WaveVision™», Rutter. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://rutter.ca/wavevision-2/>
 - [41] «Explicación de tipos de datos estructurados frente a no estructurados». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.oracle.com/es/big-data/structured-vs-unstructured-data/>
 - [42] R. KeepCoding, «5 diferencias entre datos estructurados y no estructurados». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://keepcoding.io/blog/datos-estructurados-no-estructurados/>
 - [43] «The Big Data Guide», MongoDB. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.mongodb.com/basics/big-data-explained>
 - [44] «SQL vs NoSQL: ¿Cuáles son sus diferencias? | Ventajas y Mitos». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.ilimit.com/blog/base-de-datos-sql-nosql/>
 - [45] «¿Por qué los desarrolladores prefieren las bases de datos NoSQL?» Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.oracle.com/es/database/nosql/what-is-nosql/>
 - [46] S. Martín, «Bases de datos NoSQL: Guía con las ventajas y desventajas», Pandora FMS - The Monitoring Blog. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://pandorafms.com/blog/es/bases-de-datos-nosql/>
 - [47] «The Definitive Guide to NoSQL Databases | Toptal®», Toptal Engineering Blog. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.toptal.com/database/the-definitive-guide-to-nosql-databases>
 - [48] «Base de datos NoSQL: ¿qué es NoSQL? | Microsoft Azure». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-nosql-database>
 - [49] B. Kumar, «CAP Theorem and NoSQL Databases», Medium. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://medium.com/@kumar.barmanand/cap-theorem-and-nosql-databases-589e26e15905>
 - [50] «The Big Data Guide», MongoDB. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.mongodb.com/basics/big-data-explained>
 - [51] «3 tips for building a better data lake», The Data Roundtable. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://blogs.sas.com/content/datamanagement/2019/07/22/3-tips-build-better-data-lake/>
 - [52] «Big Data Technologies: List, Stack, And Ecosystem In Demand». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.intellspot.com/big-data-technologies/>
 - [53] J. Kutay, «Data Warehouse vs. Data Lake vs. Data Lakehouse: An Overview of Three Cloud Data Storage Patterns», Striim. Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.striim.com/blog/data-warehouse-vs-data-lake-vs-data-lakehouse-an-overview/>
 - [54] D. Alzamendi, «Introduction to Data Lakehouses - David Alzamendi». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://davidalzamendi.com/intro-to-data-lakehouses/>
 - [55] «Secure satcoms: a vital capability for navies | Thales Group». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.thalesgroup.com/en/secure-satcoms-vital-capability-navies>
 - [56] «Naval SATCOM terminals | Thales Group». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.thalesgroup.com/en/markets/defence-and-security/radio-communications/naval-communications/long-range-communications-1>
 - [57] «AQUILON | Thales Group». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://www.thalesgroup.com/en/markets/defence-and-security/radio-communications/naval-communications/ship-communications/aquilon>
 - [58] «SWave. Naval communications systems | Leonardo». Accedido: 22 de diciembre de 2023. [En

línea]. Disponible en:
<https://electronics.leonardo.com/documents/16277707/18299996/Naval+Communications+Systems+mm08977.pdf?t=1669300320186>

- [59] «Hierarchical Grid Systems - Tiling Tools - Spectus - A Cuebiq Group LLC Company». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://spectus.ai/web-apps/tiling-tools/>
- [60] «Introduction | H3». Accedido: 22 de diciembre de 2023. [En línea]. Disponible en: <https://h3geo.org/docs/>
- [61] García Valera, Pablo, M. Fernández Gavilanes, y C. Pérez Collazo, «Gemelo Digital Oceanográfico: Discretización del entorno marino mediante un modelo multinivel de mallas H3», 2023.