



Escuela Técnica Superior de  
**Ingeniería Informática**

TRABAJO FIN DE GRADO

# **Despliegue de Broadsea para la estandarización con ATLAS de un estudio clínico convertido a OMOP CDM**

Realizado por  
**Da. María del Valle Alonso de Caso Ortiz**

Para la obtención del título de  
Grado en Ingeniería de la Salud

Dirigido por  
Dr. Julián Alberto García García  
Dra. María José Escalona Cuaresma

En el departamento de  
Lenguajes y Sistemas Informáticos

**Convocatoria de junio, curso 2023/24**

*A mi padre y a mi madre, por inculcarme la pasión por el estudio y acompañarme incondicionalmente en cada etapa del camino.*

# Agradecimientos

---

Nuevamente a mi familia, a mi padre Francisco José Alonso de Caso, a mi madre María del Valle Ortiz y a mis cuatro hermanos: Manuel, Ignacio, Quico y Juan Pablo; por haber sido apoyo incondicional e inspiración de los valores del trabajo, esfuerzo y sacrificio durante mis años de estudio y durante toda mi vida.

A todos mis compañeros de clase que en algún momento me han acompañado y ayudado durante el transcurso del grado de Ingeniería de la Salud y, especialmente, a aquellos que considero mis amigos y amigas, que no solo me han acompañado sino que también han amenizado este camino, llenándolo de diversión y pasión por nosotros y por estos estudios que hemos disfrutado juntos.

A todos los profesores con los que he coincidido, especialmente a Julián y María José, que además han tutelado y supervisado este Trabajo Fin de Grado.

Por último, a todos los profesionales del departamento de Innovación Tecnológica del Hospital Universitario Virgen del Rocío, que me han guiado durante el período de las prácticas curriculares, apostando por esta iniciativa y ayudándome a llevarla a cabo tutorizando y supervisando su desarrollo, especialmente a Silvia y a Carlos.

# Resumen

---

Incluya aquí un resumen de los aspectos generales de su trabajo, en español.

**Palabras clave:** Palabra clave 1, palabra clave 2, ..., palabra clave N

# **Abstract**

---

This section should contain an English version of the Spanish abstract.

**Keywords:** Keyword 1, keyword 2, ..., keyword N

# Índice general

---

<b>1 Descripción del Proyecto . . . . .</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Marco contextual . . . . .	1
1.3. Estado del Arte . . . . .	6
1.4. Motivación . . . . .	7
1.5. Estructura de la memoria . . . . .	7
<b>2 Objetivos del Proyecto . . . . .</b>	<b>9</b>
2.1. Objetivos del TFG . . . . .	9
2.2. Objetivos Personales . . . . .	9
<b>3 Gestión del Proyecto . . . . .</b>	<b>11</b>
3.1. Participantes del Proyecto . . . . .	11
3.2. Planificación temporal . . . . .	12
3.3. Planificación financiera . . . . .	12
3.4. Identificación de riesgos y planes de contingencia . . . . .	14
<b>4 Metodología . . . . .</b>	<b>15</b>
4.1. SofIA . . . . .	15
4.2. Scrum . . . . .	15
<b>5 Marco Teórico . . . . .</b>	<b>18</b>
5.1. Introducción . . . . .	18
5.2. ¿Qué es OHDSI? . . . . .	18
5.2.1. Historia . . . . .	21
5.2.2. Actualidad . . . . .	22
5.3. ¿Cómo generar evidencia? . . . . .	23
5.4. Estándares . . . . .	26
5.4.1. El Modelo de Datos Común . . . . .	27
5.4.2. El Vocabulario . . . . .	28
5.4.3. Investigación metodológica . . . . .	29
5.5. Herramientas . . . . .	32
5.5.1. ATLAS . . . . .	32
5.5.2. Otras herramientas . . . . .	36
5.6. Conclusiones . . . . .	37
<b>6 Documento de Requisitos . . . . .</b>	<b>38</b>
6.1. Introducción . . . . .	38
6.2. Requisitos Funcionales . . . . .	38
6.2.1. Diagrama de casos de uso . . . . .	38
6.2.2. Casos de uso . . . . .	38

6.3. Requisitos no Funcionales . . . . .	39
6.4. Conclusiones . . . . .	39
<b>7 Entorno de Trabajo . . . . .</b>	<b>40</b>
7.1. Introducción . . . . .	40
7.2. Arquitectura tecnológica . . . . .	41
7.2.1. Arquitectura del sistema . . . . .	41
7.2.2. Arquitectura de Broadsea . . . . .	43
7.3. Entorno tecnológico . . . . .	44
7.3.1. Google Chrome . . . . .	44
7.3.2. Docker . . . . .	44
7.3.3. PostgreSQL . . . . .	45
7.3.4. Github . . . . .	45
7.4. Conclusiones . . . . .	46
<b>8 Caso práctico . . . . .</b>	<b>47</b>
8.1. Introducción . . . . .	47
8.2. Estudio del HUVR . . . . .	47
8.3. Reproducción del estudio . . . . .	47
8.3.1. Datos . . . . .	47
8.3.2. Metodología . . . . .	47
8.3.3. Resultados . . . . .	48
8.4. Comparación de resultados . . . . .	48
8.5. Conclusiones . . . . .	48
<b>9 Resultados . . . . .</b>	<b>49</b>
9.1. Resultados . . . . .	49
9.2. Trazabilidad de objetivos . . . . .	49
9.3. Lecciones aprendidas . . . . .	49
<b>10 Conclusiones . . . . .</b>	<b>50</b>
<b>Bibliografía . . . . .</b>	<b>51</b>
<b>A Manual de ATLAS Broadsea . . . . .</b>	<b>55</b>
<b>B Glosario . . . . .</b>	<b>56</b>

# Índice de figuras

---

1.1. Esquema de contenidos de la sección . . . . .	1
4.1. Scrum. Extraída de [1] . . . . .	16
5.1. Banner de OHDSI. Extraído de web oficial [2] . . . . .	19
5.2. <i>Join the Journey</i> . Extraído del Libro de OHDSI [3] . . . . .	19
5.3. Ejemplo de la plancha. Extraído de la web oficial [2] . . . . .	20
5.4. Dibujo del proceso de extracción de evidencia. Extraído de la web oficial [2] . . . . .	21
5.5. Mapa de colaboradores de OHDSI. Extraído de la web oficial [2] . . . . .	22
5.6. Banner del Symposium Europeo 2024. Extraído de la web oficial [2] . . . . .	23
5.7. <i>The Journey from Data to Evidence</i> . Extraído del Libro de OHDSI [3] . . . . .	24
5.8. <i>The patient Journey</i> . Extraído de la página web oficial [3] . . . . .	25
5.9. Tres vías para la implementación de un análisis observacional. Extraído del Libro de OHDSI [3] . . . . .	26
5.10. Estructura del CDM v5.4. Extraída de la página de github [4] . . . . .	28
5.11. Captura de pantalla del menú principal de ATHENA . . . . .	29
5.12. Esquema simplificado de los casos de uso para la investigación en OHDSI. Extraído del Symposium 2023, publicado en la web oficial [2]	30
5.13. Esquema de los casos de uso encuadrado en la historia del paciente. Extraído del Symposium 2023 publicado en la web oficial [2] . . . . .	30
5.14. Building blocks para la definición de una cohorte. Extraído del Tutorial 2022 publicado en la web oficial [2] . . . . .	31
5.15. Logo de ATLAS. Extraída del repositorio de github [5] . . . . .	32
5.16. Biblioteca de Métodos OHDSI que utiliza ATLAS. Extraída del Libro de OHDSI [3] . . . . .	33
5.17. Estructura de la WebAPI. Extraída de la wiki de github [6] . . . . .	34
5.18. Captura de pantalla del menú principal de ATLAS demo . . . . .	35
6.1. Boceto de diagramas de casos de uso . . . . .	38
7.1. Esquema sencillo de Broadsea. Extraída de [7]. . . . .	40
7.2. Historial de versiones de Broadsea. Extraída de [7]. . . . .	41
7.3. Esquema de arquitectura <i>three-tier</i> en Docker. Extraída de [8]. . . . .	42
7.4. Vista general de todos los componentes de Broadsea. Extraída de [7].	43

# Índice de tablas

---

3.1. Descripción del primer participante del proyecto . . . . .	11
3.2. Descripción del segundo participante del proyecto . . . . .	11
3.3. Descripción del tercer participante del proyecto . . . . .	11
3.4. Descripción del cuarto participante del proyecto . . . . .	12
3.5. Descripción del quinto participante del proyecto . . . . .	12
4.1. Comparación de características entre metodologías tradicionales y ágiles en proyectos informáticos. . . . .	16

# **Índice de extractos de código**

---

# 1. Descripción del Proyecto

---

Este primer capítulo del Trabajo Fin de Grado (TFG) se divide en cinco secciones: [1.1 Introducción](#), [1.2 Marco contextual](#), [1.3 Estado del Arte](#), [1.4 Motivación](#) y [1.5 Estructura de la memoria](#).

## 1.1. Introducción

Este proyecto se desarrolla en el marco contextual del auge de la Sanidad 4.0, consecuencia de la aplicación de las nuevas tecnologías de la Industria 4.0 en el sector sanitario. Es importantísimo conocer las características de este marco contextual porque son las que definen las nuevas necesidades en el tratamiento de la información sanitaria y dan sentido a la solución que ofrece el TFG.

En las siguientes secciones se presenta el contexto teórico del problema y la solución que aporta la organización Observational Health Data Science and Informatics (OHDSI), que es la que se emplea en este trabajo para satisfacer las necesidades descritas. Además, en el estado del arte se presentan soluciones alternativas que se están llevando a cabo de forma global y europea.

Posteriormente, se presenta la motivación personal de la alumna para realizar el trabajo y la colaboración con el Hospital Universitario Virgen del Rocío en esta labor. Para acabar con se expone brevemente la estructura y los contenidos que se tratan en esta memoria del proyecto y anexos.

## 1.2. Marco contextual

El marco contextual de la problemática actual se encuentra influenciada por el surgimiento de la Industria 4.0 y las nuevas tecnologías que la acompañan, así como por su impacto significativo y transformador en el sector sanitario que ha generado nuevas necesidades a nivel mundial y europeo, más concretamente en el tratamiento de la información sanitaria. En este ámbito, la organización OHDSI se levanta como una solución innovadora y potente para paliar las necesidades de la industria.



**Figura 1.1:** Esquema de contenidos de la sección

## Industria 4.0 y nuevas tecnologías

La Industria 4.0, o cuarta revolución industrial, fue un concepto concebido por el gobierno alemán en noviembre de 2011 como una estrategia tecnológica para abordar el crecimiento industrial proyectado para 2020. Este concepto representa la cuarta fase de la industrialización, sucediendo a la mecanización, electrificación e informatización, y destaca la integración digital de tecnologías avanzadas [9].

Se centra principalmente en la digitalización y la necesaria convergencia entre los sistemas físicos y cibernéticos (*Cyber-Physical Systems, CPS*). Esta integración se busca a través de nuevas tecnologías de la información y telecomunicación (TICs), como el internet de las cosas (*Internet of Things, IoT*), la generación y análisis de datos masivos (*Big Data & Big Data Analytics*), la computación en la nube (*Cloud Computing*) y el auge de la Inteligencia Artificial (IA) [9, 10, 11]

## Características de la Sanidad 4.0

La integración de los principios y tecnologías de la Industria 4.0 en el sector sanitario originó el concepto de Salud o Sanidad 4.0 (*Healthcare 4.0*) [11, 12]. Esto origina un nuevo paradigma del que se destacan tres características principales: la provisión continua de cuidado sanitario, la orientación de la medicina hacia el paciente y la prevención y predicción de enfermedades.

- **Cuidado sanitario continuo (*continuum of care*)**. Gracias a tecnologías TIC y el IoT, la sociedad está altamente conectada, lo que ha impulsado el desarrollo de la telemedicina y la e-Salud, especialmente a raíz de la pandemia del COVID-19 [13]. Se han desarrollado numerosos dispositivos portátiles (*wearable devices*), como pulseras y relojes inteligentes, para monitorear a los pacientes tanto dentro como fuera del hospital. Además estos dispositivos generan grandes cantidades de datos médicos que se combinan con registros clínicos para formar los llamados 'Datos del mundo real' (*Real World Data, RWD*) [14].

La gran variedad en cuanto a la estructura y finalidad con la que se recopilan estos datos representan un desafío significativo en el tratamiento de la información clínica. La organización OHDSI pondrá fin a este desafío mediante la estandarización de las bases de datos RWD con un mismo fin: la investigación observacional.

- **Centrada en el paciente**. Esta perspectiva enfatiza al paciente como el eje central de la atención sanitaria [11]. Con el avance de la medicina de precisión y el seguimiento remoto de la actividad diaria, la atención médica se vuelve cada vez más personalizada [15]. La Unión Europea promueve esta orientación, exigiendo una reestructuración del sistema sanitario para que el paciente sea el principal beneficiario, evaluador y centro de los servicios de salud digital [16, 17]. Esto implica la implementación de sistemas informáticos que administren el historial clínico electrónico (HCE) completo de cada individuo, incluyendo observaciones de datos médicos, farmacéuticos y otros relevantes.

De esta forma, OHDSI presenta un modelo de datos lógico en el que el paciente es el núcleo central y alrededor de él se recoge información interseccional muy diversa, como medicamentos, procedimientos clínicos, etc.

- **Preventiva y predictiva.** Esta característica implica un enfoque proactivo en la salud en lugar de uno reactivo. La medicina se orienta hacia la prevención de enfermedades, utilizando análisis detallados del historial clínico del paciente y técnicas de aprendizaje automático (*Machine Learning, ML*) para predecir y prevenir enfermedades antes de su aparición [15]. Se emplean algoritmos avanzados de inteligencia artificial y aprendizaje automático, así como herramientas sofisticadas de análisis de datos, para abordar este desafío complejo y evolucionar hacia una atención médica más preventiva y predictiva.

La organización OHDSI presenta técnicas de ML embebidas en su herramienta de análisis por excelencia, ATLAS, que serán utilizadas en la reproducción del estudio al final del TFG.

### Necesidad de interoperabilidad

La interoperabilidad entre sistemas y datos es el objetivo principal de la actual revolución industrial, tecnológica y sanitaria. Esta necesidad es fundamental en todos los sectores y sistemas de información de organizaciones públicas y privadas, y ha sido reconocida por la Comisión Europea desde principios de siglo [18]. En 2013, el IEEE definió la interoperabilidad como "la habilidad de los sistemas de intercambiar información y utilizarla de forma efectiva".

Actualmente, el nuevo Marco de Interoperabilidad Europea (*new EIF*, 2017) se encarga de ofrecer recomendaciones para mejorar la calidad de los servicios públicos europeos en términos de interoperabilidad, ya que se considera que "la falta de interoperabilidad es el mayor obstáculo para progresar" [14]. Aunque la clasificación de los tipos de interoperabilidad aún es confusa y no existe una única clasificación concreta [19], la literatura coincide generalmente en tres tipos de interoperabilidad:

- **Interoperabilidad semántica.** La implementación de estándares o estandarización consiste principalmente en establecer acuerdos entre las grandes organizaciones de la salud para definir marcos específicos a través de los que estructurar la información clínica de manera única, reduciendo el desorden y la disparidad de los datos y permitiendo el intercambio de mensajes entre sistemas pertenecientes a distintas organizaciones. Además con los estándares nace también un concepto importante: el código abierto o *Open Source* para facilitar el acceso libre a la información y permitir consensuar un estándar común.

En este caso, OHDSI aboga por la interoperabilidad semántica aportando un estándar de datos *open-source* que utiliza la gran mayoría de los otros estándares utilizados hasta el momento, bajo la premisa "adoptar en vez de inventar" (*Adopt instead of build*).

- **Interoperabilidad técnica.** Este tipo de interoperabilidad pone el foco en la conectividad, comunicación y operación relacionadas con las entidades interactivas y los elementos de middleware. En cuanto a la autenticación y autorización, el uso de estándares técnicos, protocolos de comunicación y transporte, e interfaces entre componentes [19]. La capa técnica abarca las aplicaciones e infraestructuras que vinculan sistemas y servicios. Incluye especificaciones de interfaz, servicios de interconexión e integración de datos, presentación y intercambio de datos, y protocolos de comunicación segura [20].

Para la interoperabilidad técnica entre sus sistemas, la organización propone diversas formas de implementación de su ecosistema de herramientas, sin imponer una única tecnología sino permitiendo al usuario configurar el entorno que más conveniente le sea.

- **Interoperabilidad organizacional.** Este nivel se centra en la interoperabilidad inter e intra organizacional, en cuanto a la definición común de reglas de negocio, políticas y restricciones, alineación de procesos y las acciones necesarias para hacer que las organizaciones colaboren [21]. También se refiere a cómo los sistemas de los participantes alinean sus procesos, responsabilidades y expectativas para lograr objetivos acordados comúnmente.

OHDSI no es solo una organización científica sino una *red de colaboradores* en la que los integrantes comparten la misma misión, visión y valores, entre otras políticas.

### Desafíos en el tratamiento de los datos

A pesar de las numerosas iniciativas a nivel global y europeo, la transición hacia la interoperabilidad y estandarización sigue siendo muy desafiante debido a la complejidad y sensibilidad de los sistemas de información en salud. El manejo de datos sanitarios requiere gestiones precisas con protocolos de ciberseguridad estrictos y leyes de privacidad y confidencialidad bien definidas, lo que dificulta su implementación coordinada en diferentes regiones.

Se presentan a continuación algunos de los desafíos en el tratamiento de los datos clínicos, identificados en el Foro de Seguridad y Protección de Datos organizado por la SEIS en 2024 [22, 23].

- I. **Ciberseguridad del sistema.** La ciberseguridad de los datos clínicos representa un desafío crítico. El creciente auge de amenazas ciberneticas constantes, requiere de actualizaciones y mejoras en las medidas de protección de la información médica. Las instituciones de salud deben estar a la vanguardia en la implementación de tecnologías de seguridad robustas para salvaguardar la integridad y la confidencialidad de los datos.
- II. **Confidencialidad y privacidad.** La confidencialidad y privacidad de los datos clínicos también conforma un desafío relevante. Garantizar que solo las partes autorizadas tengan acceso a la información médica de los pacientes

requiere no solo de protocolos tecnológicos sólidos, sino también de una cultura organizacional comprometida con el cumplimiento de las regulaciones de protección de datos y la ética médica. Para ello, además se necesitan protocolos de anonimización y pseudoanonimización de las bases de datos, que garanticen la privacidad de la información personal de los pacientes.

- III. **El uso secundario.** El consentimiento para el uso secundario de datos clínicos es otro aspecto crucial a considerar. A medida que se exploran nuevas formas de aprovechar los datos para la investigación y la mejora de la atención médica, es fundamental asegurar que los pacientes comprendan y otorguen su consentimiento informado para cualquier uso adicional de su información médica, respetando siempre su autonomía y derechos individuales.
- IV. **Infraestructura tecnológica.** Por último, la infraestructura tecnológica adecuada es un requisito fundamental para el manejo eficiente de los datos clínicos. La arquitectura de los datos cada vez es más compleja y requiere infraestructuras tecnológicas muy potentes y costosas. Además, la falta de interoperabilidad entre sistemas, la obsolescencia de la tecnología y las limitaciones presupuestarias pueden obstaculizar los esfuerzos para la prestación de servicios TIC de salud.

### Propuesta de solución

Por tanto, ante las necesidades y desafíos del complejo panorama sanitario actual, este Trabajo de Fin de Grado propone a la organización **Observational Health Data Science & Informatics (OHDSI)** como la solución óptima a la interoperabilidad en estudios observacionales con datos de salud, a través del Modelo de Datos Común de OMOP y la herramienta de análisis de datos **ATLAS**.

De esta forma el proyecto pretende demostrar la utilidad y los beneficios de extraer evidencia utilizando las herramientas estandarizadas de OHDSI a través de la reproducción de un estudio clínico sobre los efectos adversos de la radioterapia en pacientes oncológicos, llevado a cabo por el Hospital Universitario Virgen del Rocío, pero utilizando ATLAS en vez de la metodología tradicional.

La relevancia de OHDSI a nivel europeo es innegable, en marzo de 2020, la red de datos y evidencia de la unión europea, EHDEN (European Health Evidence & Data Network) comenzó a colaborar con OHDSI para poner fin a la disparidad de estándares presente en los distintos nodos de la unión europea y proporcionar un Modelo de Datos Común y un espacio de datos interoperable y estandarizado para todos. A partir de entonces OHDSI ha comenzado a ganar gran relevancia a través de su participación en proyectos europeos como DARWIN EU (*Data Analysis and Real World Interrogation Network European Unión*, 2022) [24] o EUCAIM (*EUropean Cancer Image*, 2023).

Además a nivel estatal, España conforma uno de los nodos de colaboración con OHDSI más grandes de Europa. y en Sevilla, especialmente, la colaboración con OHDSI la llevan a cabo el IBIS (Instituto de Biomedicina de Sevilla), la fundación

FISEVI (Fundación para la Gestión de la Investigación en Salud en Sevilla) y los hospitales universitarios Virgen Macarena y Virgen del Rocío. Este último además es la sede del convenio de prácticas en empresa realizadas junto al TFG .

### 1.3. Estado del Arte

Boceto de contenidos básicos. Más ideas???? Más temáticas?? Mayor extensión???

Frente a todo lo expuesto, aún no existe un consenso entre las grandes potencias que establezca verdaderamente una solución única. En el ámbito del tratamiento de la información sanitaria, existen numerosas alternativas a OHDSI y organizaciones proveedoras de estándares y herramientas para paliar las necesidades y dificultades del sector. Aunque irónicamente la presencia de tantísimas alternativas diferentes es la principal dificultad para la interoperabilidad.

En el ámbito de la interoperabilidad semántica, algunos de los estándares más reconocidos y usados mundialmente son HL7 FHIR (*Health Level Seven - Fast Health Interoperability Resources*), HL7 CDA (*Health Level Seven Clinical Document Architecture*), DICOM (*Digital Imaging and Communications in Medicine*), SNOMED CT (*Systematized Nomenclature of Medicine - Clinical Terms*), IHE (Integrating the Healthcare Enterprise), openEHR (Open Electronic Health Record), LOINC (Logical Observation Identifiers Names and Codes), RxNorm (Prescription Norm) entre otros.

Además, solo en España, cada comunidad autónoma utiliza un sistema informático sanitario distinto, cuyos datos están estructurados de formas distintas. En Andalucía el sistema de información es DIRAYA. Otros ejemplos son: en Madrid, Historia Clínica Digital de Atención Primaria (HCDSAP); en Cataluña, Sistema de Información de Atención Primaria (SIAP); en la Comunidad Valenciana, Sistema de Información Poblacional de Atención Primaria (SIPAP), en País Vasco, Osabide; en Galicia, SERGAS; entre otros.

Por otro lado, en el ámbito de la interoperabilidad técnica las alternativas son muy dispares, desde aquellos que realizan análisis totalmente personalizados mediante scripts de código o el gran catálogo de software de procesamiento de datos actualmente disponible en el mercado. Los lenguajes de programación que más usan los analistas de datos son Python, R y SQL, implementados en diferentes IDEs como JupyterLab o Jupyter Notebook para Python, Rstudio para R o multitud de plataformas de bases de datos (Oracle, Postgre, BigQuery...). Por otra parte, los software de análisis más extendidos son Tableau, Microsoft PowerBI, SAS, MatLab, Apache Spark, entre otras.

La falta de un estandar común es objeto de investigación en todo el mundo, lo que da lugar a alianzas entre organizaciones y competiciones en proyectos que pretenden dar solución a este aspecto, como por ejemplo la Infraestructura de Servicios Digitales de eSalud (eHDSI) [25] o el proyecto European Genomic Data Infrastructure (GDI) [26] que busca establecer una infraestructura unificada para gestionar y compartir datos genómicos en Europa.

## 1.4. Motivación

La principal motivación para realizar este proyecto ha sido mi curiosidad e interés por el mundo de la ciencia de datos a lo largo de mis años de formación universitaria. El origen se sitúa en el primer año de carrera, allá por el 2020, cuando por primera vez oí hablar del análisis de datos clínicos como una disciplina emergente de gran interés a nivel laboral. A partir de este momento, continué investigando sobre todo lo relacionado a esta disciplina y cuando en tercero de carrera tuve la oportunidad de realizar el programa de movilidad ERASMUS al Politecnico di Milano, no dudé en seleccionar el mayor número de asignaturas de Data Science que mi convenio de estudios me permitió.

Este año de estudio en Milán confirmó que, lo que había nacido como una mera curiosidad, se había convertido en una pasión, por lo que a mi regreso del Erasmus me decidí a orientar mi carrera profesional y mi TFG hacia el mundo del análisis de datos clínicos, hasta el día de hoy en que este Trabajo Fin de Grado es escrito.

El Trabajo Fin de Grado ha sido realizado por mí, María del Valle Alonso de Caso Ortiz, alumna del grado de Ingeniería de la Salud por la Universidad de Sevilla (US), de la promoción 2020-2024 y bajo la tutela de D. Julián A. García García y Da. Maria J. Escalona Cuaresma, ambos pertenecientes al departamento de Lenguajes y Sistemas Informáticos de la Escuela Técnica Superior de Ingeniería Informática (ETSII) de la misma universidad.

Además se ha realizado en conjunto con el Departamento de Innovación Tecnológica del Hospital Universitario Virgen del Rocío, mediante un convenio de prácticas curriculares de 337 horas, donde han ejercido la tutela D. Silvia Rodríguez Mejías y D. Carlos Luis Parra Calderón.

De esta forma, también ha sido de gran importancia la motivación de mis profesores y tutores de la ETSII y compañeros del grupo científico del Departamento de Innovación Tecnológica del hospital, quienes confiando en mi me han apoyado, motivado y guiado durante mi formación sobre ATLAS y OHDSI y la informática clínica en general.

## 1.5. Estructura de la memoria

La memoria se estructura en diez capítulos y dos anexos que contienen toda la información relevante.

La información propiamente sobre el proyecto se encuentra en los capítulos: [1 Descripción del Proyecto](#) [2 Objetivos del Proyecto](#), [3 Gestión del Proyecto](#) y [4 Metodología](#).

A continuación, en [5 Marco Teórico](#), se recopila información relevante sobre la organización Observational Health Data Science and Informatics (OHDSI) y sus herramientas, prestando especial atención a ATLAS. Este capítulo en conjunto con

6 Documento de Requisitos y 7 Entorno de Trabajo proveen un conocimiento teórico completo de las herramientas y análisis a tratar durante el proyecto.

Por otra parte, en 8 Caso práctico se presenta un capítulo dedicado a la reproducción de un caso práctico de estudio utilizando ATLAS.

Por último, los siguientes dos capítulos 9 Resultados y 10 Conclusiones, presentan una recopilación de resultados y conclusiones, respectivamente obtenidos al término del desarrollo del TFG.

Adicionalmente, se adjuntan dos anexos. El anexo A consiste en una guía completa de Instalación, despliegue y configuración de Broadsea y el Anexo B, se trata Glosario de Términos técnicos relevantes para la comprensión del trabajo.

Por su naturaleza informática, este TFG se ha desarrollado paralelamente a un repositorio de github del proyecto [27], que ha servido como controlador de versiones y como administrador de archivos en la nube, permitiendo almacenar y compartir con el lector archivos relevantes del TFG, ya sean archivos necesarios para el despliegue de la herramienta, archivos producidos durante el análisis o los propios documentos en sí mismos.

## **2. Objetivos del Proyecto**

---

En este capítulo se presentan los objetivos del Trabajo Fin de Grado, consensuados por el alumno, los tutores de la Universidad de Sevilla y los del Hospital Universitario Virgen del Rocío. Se presentan en [2.1](#) los objetivos generales para el desarrollo del TFG y en [2.2](#) los objetivos personales del alumno.

### **2.1. Objetivos del TFG**

Los objetivos relativos al desarrollo teórico y práctico del TFG son tres y se presentan a continuación:

1. **Obj-001: Estudio teórico de organización OHDSI y herramienta ATLAS Broadsea.** Este objetivo proporciona al alumno un marco de fundamentación y comprensión necesario para poder extraer verdadero valor del uso de ATLAS y de todo el ecosistema de la comunidad científica de OHDSI.
2. **Obj-002: Instalación, configuración y despliegue de ATLAS Broadsea.** Este objetivo, acompañado de la redacción del Anexo A, es de importancia trascendental, puesto que el anexo reúne en un único documento inédito información difícilmente accesible y desperdigada en la red, constituyendo un documento de gran relevancia para toda la comunidad científica, especialmente para el equipo del Hospital, que contará con mayor facilidad a la hora de realizar estas tareas sobre Broadsea.
3. **Obj-003: Reproducción de caso práctico de análisis de datos clínicos proporcionados por el hospital.** Este objetivo está ligado en igual medida al TFG y a las prácticas realizadas en el Hospital, pues consiste en reproducir un estudio ya realizado previamente sobre unos datos proporcionados por el HUVR pero utilizando, en este caso, ATLAS. La colaboración con el hospital en este caso es crucial para el alcance de este objetivo que de forma práctica complementa a la documentación teórica del TFG.

### **2.2. Objetivos Personales**

Los objetivos personales, relativos a la ambición, interés y curiosidad de la alumna son tres y se presentan a continuación:

1. **Obj-Pers-001: Aumentar mi conocimiento de la comunidad OHDSI y sus herramientas.** Este objetivo se debe a que inicialmente mi desconocimiento sobre OHDSI era absoluto. Por tanto, aumentar mi conocimiento sobre la organización es importante para comprender la utilidad de la misma y de las

herramientas que proporciona y poder realizar un trabajo coherente y bien fundamentado.

2. **Obj-Pers-002: Aumentar mi conocimiento del mundo del análisis de datos.** Este objetivo se debe a que, aunque es cierto que durante mis estudios de grado he aprendido y obtenido grandes conocimientos sobre las ciencias de datos, de este trabajo final también se espera aumentar en mayor profundidad los conocimientos teóricos, generales y específicos a una herramienta de gran interés europeo como es ATLAS para el análisis de datos.
3. **Obj-Pers-003: Aumentar mi experiencia laboral analizando datos clínicos.** Este objetivo pretende aumentar la experiencia de analizar datos clínicos fuera del marco meramente académico, sino en un entorno de trabajo real, con datos clínicos reales, gracias a la colaboración con el grupo de Innovación Tecnológica del Hospital.

# 3. Gestión del Proyecto

---

En este capítulo se presenta toda la información relacionada con la gestión del proyecto de la elaboración del TFG. El capítulo se divide en cuatro secciones: [3.1 Participantes del Proyecto](#), [3.2 Planificación temporal](#), [3.3 Evaluación de costes](#) y [3.4 Identificación de riesgos y planes de contingencia](#).

## 3.1. Participantes del Proyecto

Los participantes del proyecto TFG se presentan a continuación mediante una tabla que recoge su nombre, institución a la que pertenece, rol asignado durante la elaboración del proyecto e información de contacto.

Es importante destacar que los tres primeros participantes corresponden a alumna y tutores de la Escuela Técnica Superior de Ingeniería Informática de la Universidad de Sevilla y los dos últimos participantes, a los tutores de las prácticas realizadas en el Departamento de Innovación Tecnológica del Hospital Universitario Virgen del Rocío.

<b>Participante</b>	María del Valle Alonso de Caso Ortiz
<b>Institución</b>	Universidad de Sevilla
<b>Rol</b>	Jefe de Proyecto & Developer & Analista
<b>Información de contacto</b>	vallealonsodecaso@icloud.com

**Tabla 3.1:** Descripción del primer participante del proyecto

<b>Participante</b>	Julián García García
<b>Institución</b>	Universidad de Sevilla
<b>Rol</b>	Tutor del TFG & Supervisor
<b>Información de contacto</b>	juliangg@us.es

**Tabla 3.2:** Descripción del segundo participante del proyecto

<b>Participante</b>	María José Escalona Cuaresma
<b>Institución</b>	Universidad de Sevilla
<b>Rol</b>	Tutor del TFG & Supervisor
<b>Información de contacto</b>	mjescalona@us.es

**Tabla 3.3:** Descripción del tercer participante del proyecto

<b>Participante</b>	Silvia Rodríguez Mejías
<b>Institución</b>	Hospital Universitario Virgen del Rocío
<b>Rol</b>	Tutor de prácticas en empresa
<b>Información de contacto</b>	

**Tabla 3.4:** Descripción del cuarto participante del proyecto

<b>Participante</b>	Carlos Luis Parra Calderón
<b>Institución</b>	Hospital Universitario Virgen del Rocío
<b>Rol</b>	Supervisor de prácticas en empresa
<b>Información de contacto</b>	

**Tabla 3.5:** Descripción del quinto participante del proyecto

## 3.2. Planificación temporal

- Scrum, planificación por sprints, estimación del tiempo, desviación...

## 3.3. Planificación financiera

La planificación financiera se realiza de forma similar a la elaboración de un presupuesto sobre el proyecto. Para ello se realizará el cálculo de dos tipos de coste: personal y material. Por último se estimará el coste total y el beneficio.

### Coste de personal

Para el coste de personal se tendrán en cuenta los roles definidos previamente (véase [3.1](#)). Concretamente, intervendrán los roles ejercidos por la alumna y se omitirán los roles de tutorización y supervisaje para el cómputo del presupuesto del proyecto.

Por tanto, el proyecto requiere del ejercicio fundamental de tres roles: jefe de proyecto, developer y analista. El jefe de proyecto asume las tareas de comunicarse con la empresa externa, tomar decisiones y acordar objetivos y elaboración del trabajo. El developer asume las tareas de instalar, desplegar y configurar el sistema así como gestionar y administrar las bases de datos y asegurar el correcto funcionamiento de la herramienta. Por último, el analista realiza las tareas meramente analíticas, se encarga de la reproducción del estudio clínico *per se* haciendo uso de la herramienta una vez instalada.

Los costos de cada rol se calculan por hora, utilizando como referencia el precio medio publicado en la consulta preliminar para perfiles profesionales del ámbito informático [28], considerando la categoría junior para cada uno.

A continuación se presenta en negrita el rol definido en el proyecto seguido de la categoría a la que se ha asociado el rol según el informe de la Junta y el coste total asociado a las horas invertidas.

- **Jefe de proyecto.** Jefe de proyecto y coordinador junior: 39.16€/h.

$$39,16\text{ €/h} \times 700\text{ h} = 27412\text{ €} \quad (3.1)$$

- **Developer.** Administrador de la base de datos junior: 35.18€/h.

$$35,18\text{ €/h} \times 700\text{ h} = 27412\text{ €} \quad (3.2)$$

- **Analista.** Analista funcional de aplicaciones junior: 33.12€/h

$$33,12\text{ €/h} \times 700\text{ h} = 27412\text{ €} \quad (3.3)$$

Por tanto, el **coste total destinado al personal del proyecto es XXXX €.**

### Coste material

En cuanto a los costes materiales, se distinguen otras tres categorías: costes de amortizaciones, de licencias y de servicios.

#### cambiar datos

En primer lugar, el coste de amortizaciones tendrá en cuenta únicamente el equipo portátil utilizado para el desarrollo del proyecto. Se realizará una amortización lineal en 5 años, con un coste inicial de 1000 € y un valor residual del 20 de este coste inicial que da lugar a un coste de 13.33€/mes.

- **Equipo portátil.** Ordenador con procesador 7th generation y 8 gb de RAM.

$$\text{valor residual} = 1000\text{ €} \times 0,20 = 200\text{ €} \quad (3.4)$$

$$\text{valor total} = \frac{1000 - 200}{60 \text{ meses}} = 13,33\text{ €/mes} \quad (3.5)$$

En segundo lugar, el coste de licencias tendrá en cuenta el uso de software de pago. La mayoría de las herramientas utilizadas durante el proyecto poseen un plan gratuito o son gratuitas en sí mismas, a excepción de las siguientes:

- **Licencia de Windows 11 Pro [29]:** 259€
- **Licencia profesional de Enterprise Architect [30]:** 229€
- **Licencia de Microsoft Office 365 [31]:** 69€

Tenemos por tanto un coste total de licencias de 557€.

En tercer lugar, los costes de servicios incluyen los gastos por suministro eléctrico, el cual tiene un coste promedio de 0,182 € / KWh (OCU, 2022) . Se estima un consumo medio de 0,3 KWh de los dispositivos electrónicos usados durante el desarrollo.

■ **Suministro eléctrico.**

$$0,182 \text{ €/kWh} \times 0,3 \text{ kWh} \times 300 \text{ h} = 16,38 \text{ €.} \quad (3.6)$$

También es necesario tener en cuenta el servicio de internet, para el que se tiene contratado un servicio de fibra óptica simétrica de 100 megabytes con un coste mensual de 25,70 €:

■ **Suministro de internet.**

$$25,70 \text{ €/mes} \times 4 \text{ meses} = 102,8 \text{ €.} \quad (3.7)$$

Con esto obtenemos un **coste material total de 727,38€.**

### **Coste total y beneficio**

#### **cambiar datos**

Sumando los costes de personal y materiales, se tiene que el coste total estimado del proyecto asciende a la cifra de 8.959,42 €.

En cuanto al beneficio que tiene este proyecto, ha sido estimado como el coste total más un 15 % de beneficio íntegro para la empresa. Esto hace un total de 10.303,33 € de beneficio total del proyecto.

Por último, se añadirá un fondo de contingencia frente a riesgos del proyecto que quedará excluido de este coste total. Su finalidad será evitar que alguna desviación o problema pueda provocar una finalización temprana del proyecto. Correspondrá al 10 % del coste total estimado, por lo que se contará con un fondo de contingencia de 895,94 €.

### **Desviaciones**

#### **cambiar datos**

## **3.4. Identificación de riesgos y planes de contingencia**

- Quedarme sin wifi para trabajar en latex
- Que se caiga el servidor de latex

# **4. Metodología**

---

A continuación se presentan las metodologías empleadas para desarrollar el proyecto. Para la redacción de la documentación y catálogo de requisitos se ha utilizado el software [4.1 SofIA](#) y para la planificación temporal se ha utilizado la metodología de gestión de proyectos [4.2 Scrum](#)

## **4.1. SofIA**

SofIA [32] es una metodología web dirigida por modelos, cuyo propósito inicial consistió en brindar respaldo a los requisitos del desarrollo web. Fue implementada por primera vez como NDT aunque actualmente ha evolucionado para ofrecer soporte a todo el ciclo de desarrollo, abarcando fases como estudio de viabilidad, requisitos, análisis, diseño, implementación, pruebas y mantenimiento.

Este proyecto se ha beneficiado enormemente del uso de SofIA especialmente en la fase de requisitos, que es el núcleo de esta metodología y la razón principal para seguir sus técnicas. Estas facilitarán la captura, definición y validación de una amplia variedad de requisitos.

SofIA no solo ofrece técnicas tradicionales como trazabilidad o prototipos, sino que también aborda otros aspectos como la navegación entre componentes. Esto garantiza una conexión entre todos los elementos y evita inconsistencias en el catálogo de requisitos después de una modificación.

Es importante destacar que SofIA cuenta con un alto grado de automatización y se basa en la herramienta profesional Enterprise Architect [30]. En los últimos años, la metodología NDT ha sido ampliamente utilizada como enfoque principal en numerosos proyectos reales de importantes compañías. Entre ellas se destacan entidades públicas como la Consejería de Salud de Andalucía o la Consejería de Cultura de la Junta de Andalucía, así como empresas privadas como Airbus o Everis.

Esta amplia adopción refleja un alto grado de confianza en la metodología, y se garantiza que su aplicación conducirá al proyecto a desarrollarse en un entorno comparable al de cualquier otro proyecto real, como es el caso de este proyecto.

## **4.2. Scrum**

Scrum [1] consiste en una metodología ágil para la gestión y planificación de proyectos informáticos. En el campo de la informática las metodologías ágiles están en auge, cada vez se opta menos por metodologías tradicionales y se apuesta por estas nuevas metodologías disruptivas. Los motivos y beneficios de ello son muy

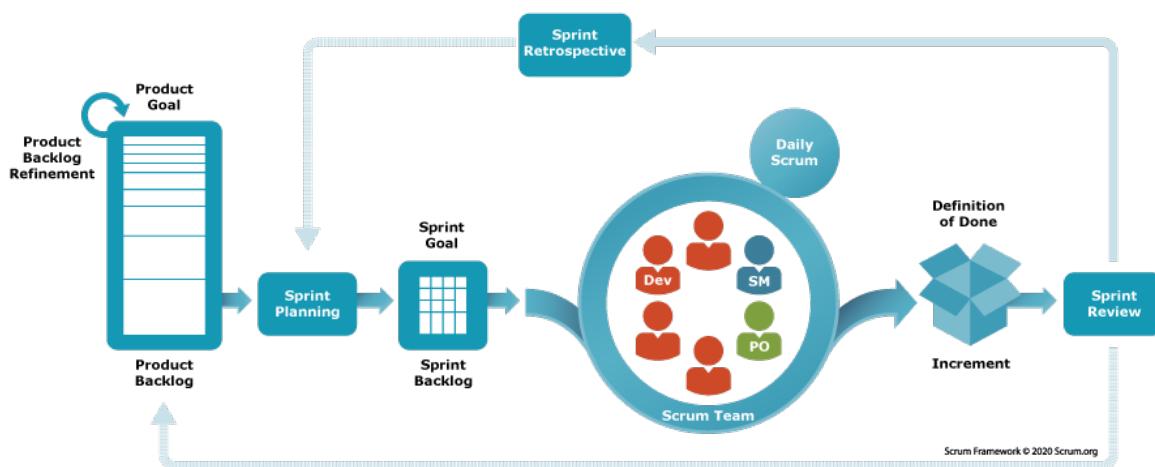
numerosos, las metodologías ágiles abogan por el cambio continuo y la adaptabilidad frente a la rigurosidad tradicional.

A continuación se presenta una tabla esquemática de beneficios en el uso de metodologías ágiles.

Característica	Metodologías Tradicionales	Metodologías Ágiles
Planificación	Planificación detallada y rígida al inicio del proyecto.	Planificación adaptable y flexible, se adapta a cambios constantes.
Entrega de valor	Entregas al final del proyecto.	Entregas frecuentes de funcionalidades, permitiendo feedback temprano.
Cambio	Cambios difíciles de gestionar, conllevan retrasos y costos adicionales.	Cambios bienvenidos y gestionados de manera eficiente, se incorporan fácilmente al proyecto.
Cliente	Interacción limitada con el cliente.	Colaboración estrecha con el cliente, involucrado en todo el proceso.

**Tabla 4.1:** Comparación de características entre metodologías tradicionales y ágiles en proyectos informáticos.

Entre las diversas metodologías ágiles, concretamente se ha seleccionado Scrum, que es una solución que se basa en numerosos ciclos iterativos, denominados *sprints*, para el desarrollo incremental del producto final.



**Figura 4.1:** Scrum. Extraída de [1]

En la Figura 4.1 se presentan algunos elementos importantes intrínsecos de la

metodología de Scrum que han intervenido en la gestión del proyecto, como son:

- **Daily Standup:** Reuniones diarias para el seguimiento del proyecto, donde se exponen los avances y problemas o dificultades que se han tenido en el transcurso del ciclo diario del proyecto. En el proyecto, esta práctica se llevó a cabo a través de la monitorización continua realizada por los tutores del HUVR.
- **Product Owner:** Rol responsable de las características del producto y de asegurar que el equipo aporte valor a la empresa. En el proyecto, este rol es desempeñado por el tutor del HUVR, Carlos Parra (véase [3.1](#)).
- **Product Backlog:** Lista priorizada de características que debe tener el producto a desarrollar. En el proyecto, este elemento corresponde al catálogo de requisitos definido conjuntamente con los tutores de la universidad, Julián García y María José Escalona (véase [3.1](#)).
- **Sprint Backlog:** Conjunto de características escogidas del Product Backlog para implementar en el sprint. En el proyecto, en cada sprint se realizó una división de tareas a realizar, y dentro de cada una se incluyeron un conjunto de subtareas. En esta selección de tareas interviene mayoritariamente la alumna, que es Jefe del Proyecto.
- **Sprint Review:** Reunión en la que se presenta y se evalúa el trabajo realizado durante el sprint, con el objetivo de conseguir la aprobación por parte del cliente. En el proyecto, se considera el cliente a los tutores de la universidad, que son los evaluadores del proyecto.

El resto de los elementos que aparecen en la Figura [4.1](#) y no se han definido anteriormente, se debe a que no han tenido una aplicación práctica real en el transcurso del proyecto.

# 5. Marco Teórico

---

Este capítulo presenta el marco teórico sobre OHDSI: [5.1](#) Introducción breve al capítulo, [5.2](#) Qué es OHDSI, [5.3](#) Cómo generar evidencia, [5.5](#) Qué herramientas ofrece y [5.6](#) Breve conclusión final.

## 5.1. Introducción

La organización Observational Health Data Science and Informatics (OHDSI) es muy importante para el TFG por ser la proveedora de la herramienta de análisis ATLAS, núcleo central del trabajo, y por la relevancia que ha adquirido a nivel europeo en los últimos años.

Por tanto, en este capítulo se pretende dar a conocer al lector el ecosistema completo de la organización que provee la herramienta, en cuanto a los aspectos más generales como los más específicos. No se puede entender ATLAS sin entender OHDSI porque las herramientas no están aisladas sino que forman parte de un ecosistema dependiente entre sí. Todas estas dependencias y conceptos necesarios se exponen en este capítulo para generar un conocimiento profundo subyacente al caso práctico (véase [8](#)) y para poder volver a estos conceptos básicos cuando fuere necesario.

## 5.2. ¿Qué es OHDSI?

OHDSI, pronunciado en inglés "Odysee", son las siglas de *Observational Health Data Science and Informatics*. OHDSI es una organización colaborativa de ciencia abierta cuyo propósito, de forma muy resumida, es mejorar la investigación científico-sanitaria a través de la ciencia de datos y la informática clínica. No obstante, no es solo una organización, sino una comunidad global abierta a todo el que esté interesado y alineado con su misión, visión y objetivos.

La comunidad se asigna por tanto la misión de "mejorar la salud empoderando a una comunidad para generar de manera colaborativa evidencia que promueva mejores decisiones de salud y una mejor atención", y comparte la visión de "un mundo en el que la investigación observacional produzca una comprensión integral de la salud y la enfermedad" [\[2\]](#)[\[3\]](#).

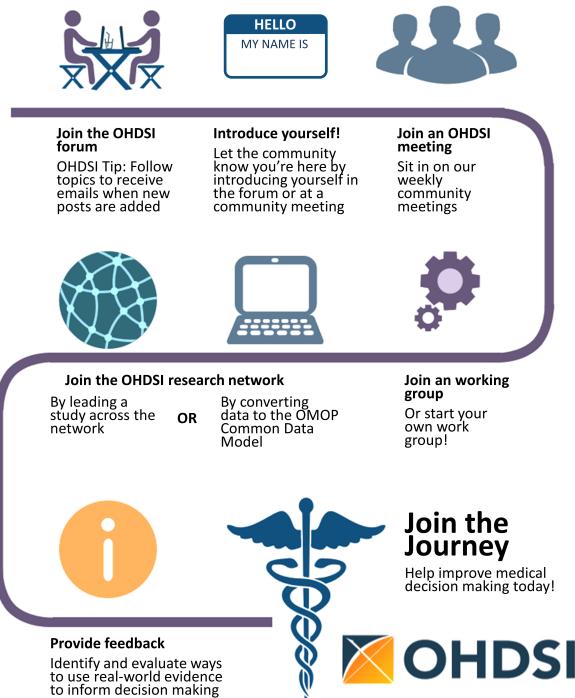
Por otra parte, en El Libro de OHDSI la organización se define así misma como "una comunidad de ciencia abierta que tiene como objetivo mejorar la salud empoderando a la comunidad para generar de manera colaborativa evidencia que promueva mejores decisiones de salud y mejor atención" [\[3\]](#).



**Figura 5.1:** Banner de OHDSI. Extraído de web oficial [2]

Por tanto, a la pregunta sobre *qué es OHDSI* se puede responder apoyándose en cuatro características fundamentales: (i) una comunidad o red colaborativa, (ii) de ciencia abierta, (iii) estandarizada y (iv) con la finalidad de promover la extracción de evidencia a partir de datos clínicos.

- I. **Una comunidad o red colaborativa.** La organización es una comunidad, es decir, se presenta abierta a la incorporación de todo aquel que esté comprometido con su misión. Además se muestra siempre abierta e interesada en la incorporación de nuevos colaboradores, lo que muestran constantemente con el eslogan “Join the Journey”, en español, “únete a la aventura” (véase Figura 5.2). Además la organización distribuye a sus colaboradores en nodos por países y en grupos de trabajo según los diferentes componentes de OHDSI.



**Figura 5.2:** *Join the Journey*. Extraído del Libro de OHDSI [3]

- II. **Ciencia abierta.** La forma de trabajar de la organización es muy importante, puesto que promueve la colaboración entre las organizaciones y la

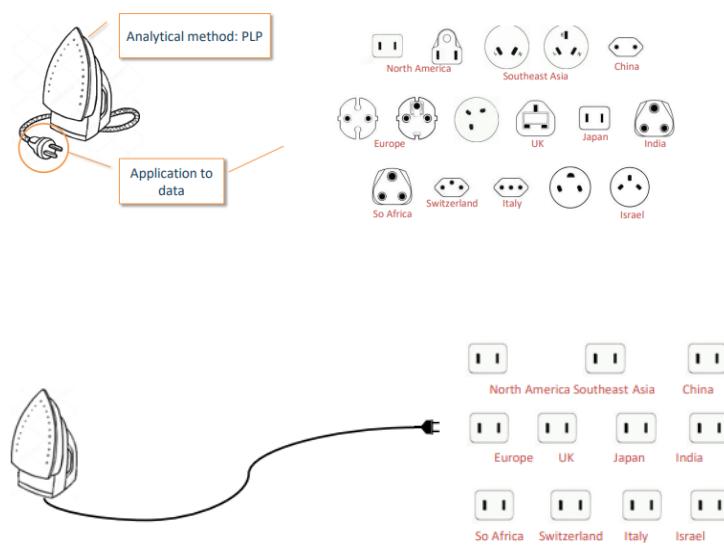
participación a través de la ciencia abierta.

Todos los eventos, publicaciones, herramientas y documentación que elabora OHDSI están disponibles públicamente y de forma gratuita en internet, para que pueda unirse quien quiera (en el caso de los eventos) o consultarse y usarse en cualquier momento (en caso de las herramientas e información). Las dos vías de información por excelencia sobre OHDSI son su página web [2] y el *Libro de OHDSI* [3]. Otras vías de información son a través de publicaciones científicas [33], tutoriales para principiantes, grabaciones de las reuniones semanales de la comunidad o las conferencias anuales a través de su canal de youtube [34], canales de mensajería abierta como discord [35] o MS Teams [36], cientos de repositorios de github con información técnica de cada herramienta [37] y los foros de la comunidad para solventar dudas y preguntas [38], entre otros.

Además, OHDSI asegura la fiabilidad, accesibilidad, interoperabilidad y reproducibilidad de sus estudios a través del cumplimiento de los principios FAIR, que desarrolla en gran extensión en la sección 3.7 de su libro [3].

**III. Estandarización.** OHDSI aboga por estandarizar los modelos de datos y la metodología de la investigación médica a un modelo común, con la finalidad de aumentar la interoperabilidad entre los sistemas y organizaciones sanitarias a nivel mundial.

Esta idea se presenta en el Symposium de 2023 con un ejemplo muy intuitivo: la conexión a la corriente eléctrica a través de una plancha. La conexión de la plancha sería la realización de un estudio sobre unos datos, que serían el enchufe a la corriente eléctrica, siendo el objetivo establecer un enchufe estándar que permita la conexión de la plancha a la corriente eléctrica en cualquier lugar del mundo, es decir, la realización de un estudio siguiendo una misma estructura en cualquier lugar del mundo.



**Figura 5.3:** Ejemplo de la plancha. Extraído de la web oficial [2]

**IV. Extracción de evidencia a partir de datos clínicos.** Es importante destacar la finalidad de OHDSI de, no solo recopilar y almacenar la información clínica, sino también extraer información o evidencia de ella. Este es el propósito final y enfrenta muchos desafíos debido a la disparidad de los datos y técnicas de análisis.

El proceso de extracción de evidencia no es sencillo, como se muestra en la Figura 5.4, y parte en un extremo de las diferentes bases de datos del mundo real (RWD) (véase 1.2) hacia la obtención fiable de evidencia del mundo real (RWE). Este recorrido debe atravesar un proceso de ETL para estandarizar las bases de datos al modelo común de OMOP, la integración con el vocabulario y el análisis técnico en sí que permita extraer la evidencia. Por suerte, la organización también proporciona un conjunto de herramientas (véase 5.5) para realizar más sencillamente todo el recorrido.



**Figura 5.4:** Dibujo del proceso de extracción de evidencia. Extraído de la web oficial [2]

### 5.2.1. Historia

Es común encontrar en internet los términos OHDSI y OMOP (*Observational Medical Outcomes Partnership*), utilizados de forma casi indistintiva. Si bien es verdad que OMOP se suele asociar mayoritariamente al CDM (*Common Data Model*) también OHDSI mantiene gran relación con este modelo común de datos. Entonces, ¿cuál es la relación entre estas dos entidades? Pues bien, la iniciativa de OHDSI se originó en 2014, posterior al proyecto OMOP, que finalizó en 2013, pues la relación que guardan estas dos entidades es filial, OHDSI es la sucesora de OMOP.

OMOP nació en 2008 como una asociación público-privada presidida por la Administración de Alimentos y Medicamentos de EE. UU. con el objetivo de establecer buenas prácticas en estudios observacionales retrospectivos. El proyecto además fue administrado por la Fundación de los Institutos Nacionales de Salud y financiado por un consorcio de compañías farmacéuticas en colaboración con otros investigadores académicos y socios de datos de salud [39]. El propósito inicial de OMOP era impulsar la ciencia de la vigilancia activa de la seguridad de los productos médicos mediante el análisis de datos observacionales de atención médica [39]. Sin embargo, durante su desarrollo, se enfrentó a los desafíos técnicos

de llevar a cabo investigaciones en bases de datos observacionales muy heterogéneas entre sí.

El resultado fue el desarrollo de un Modelo Común de Datos (CDM) como un mecanismo para estandarizar la estructura, el contenido y la semántica de los datos observacionales y hacer posible escribir código de análisis estadístico que fuera reutilizable para estudios en distintas fuentes de datos [40]. Los experimentos de OMOP demostraron la viabilidad de establecer un CDM que además reuniese diferentes vocabularios estandarizados, reuniendo en un mismo estándar diversos tipos de datos de diferentes entornos de atención y representados por diferentes vocabularios de origen. Esta característica facilitó la colaboración y aumentó el interés entre diferentes instituciones lo que promovió un enfoque de ciencia abierta [3]. OMOP puso todo su trabajo a disposición del público, incluidos diseños de estudio, estándares de datos, código de análisis y hallazgos empíricos, para mejorar la transparencia y fomentar la confianza en su investigación.

Al término del proyecto, el Modelo Común de Datos (CDM) de OMOP había evolucionado hasta respaldar un abanico amplísimo de aplicaciones analíticas, incluida la efectividad comparativa de intervenciones médicas y políticas de todo el sistema de salud, no solo de la industria farmacéutica, por tanto, el equipo de investigación acordó que el fin de dicho proyecto debería ser el origen de uno nuevo. A partir de esta idea nació OHDSI [3].

### 5.2.2. Actualidad

Por tanto, lo que nació en 2014 como la continuación del proyecto OMOP ha evolucionado hasta convertirse en una extensa red colaborativa global. En la actualidad, la comunidad de OHDSI cuenta con la participación de más de tres mil colaboradores distribuidos en 80 países.



**Figura 5.5:** Mapa de colaboradores de OHDSI. Extraído de la web oficial [2]

Además, tal y como se presenta en [1.3](#), desde que se inició su colaboración con EHDEN (European Health Data Evidence) en 2020, OHDSI está adquiriendo cada vez mayor relevancia a nivel europeo. Ejemplo de ello es la celebración, este mes de junio, en Rotterdam del quinto Symposium Europeo de OHDSI (véase Figura [5.6](#)), que tiene el fin de reunir a los expertos y miembros de la comunidad para presentar los grandes proyectos nacionales y europeos que se están realizando en toda europa con las herramientas de la comunidad.



**Figura 5.6:** Banner del Symposium Europeo 2024. Extraído de la web oficial [\[2\]](#)

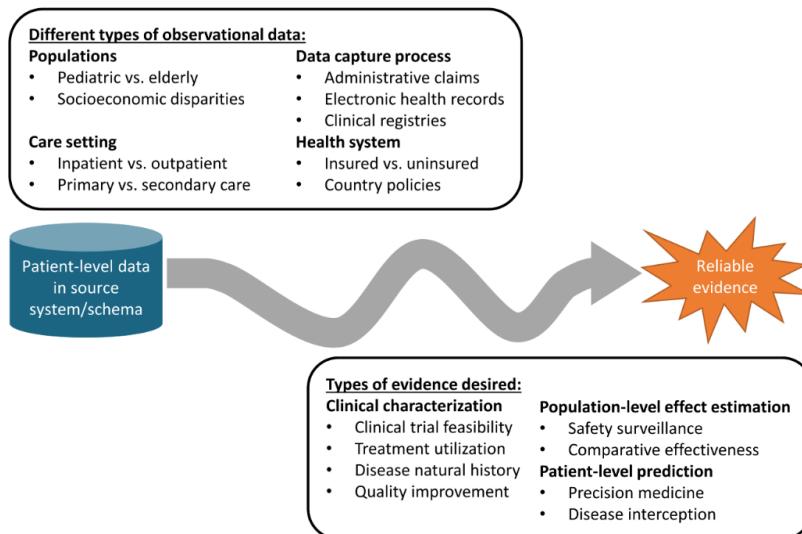
Por ejemplo, en el Symposium Europeo del pasado año 2023, se presentaron proyectos relativos al almacenamiento de los datos de UCI en Holanda [\[41\]](#), la integración del CDM de OMOP con el laboratorio de datos de salud alemán [\[42\]](#), la estandarización de la base de datos nacional francesa SNDS al modelo de OMOP [\[43\]](#), la armonización de los HCE hospitalarios en Ruanda al CDM [\[44\]](#) y a la estandarización de los datos del registro europeo de sarcomas a OMOP [\[45\]](#), entre otros.

### 5.3. ¿Cómo generar evidencia?

Una vez que se conoce qué es OHDSI su misión y sus características fundamentales, se conoce la importancia de generar evidencia a partir del estudio de los datos clínicos. A continuación se exponen los principios fundamentales de la organización para generar evidencia confiable.

#### The journey from data to evidence

No es casualidad que la invitación que hace OHDSI a sus colaboradores lleve el slogan "*Join the Journey*" sino que es un guiño al propósito al que se unen, es decir, al camino desde los datos hasta la evidencia o, en inglés, "*The Journey from data to evidence*".

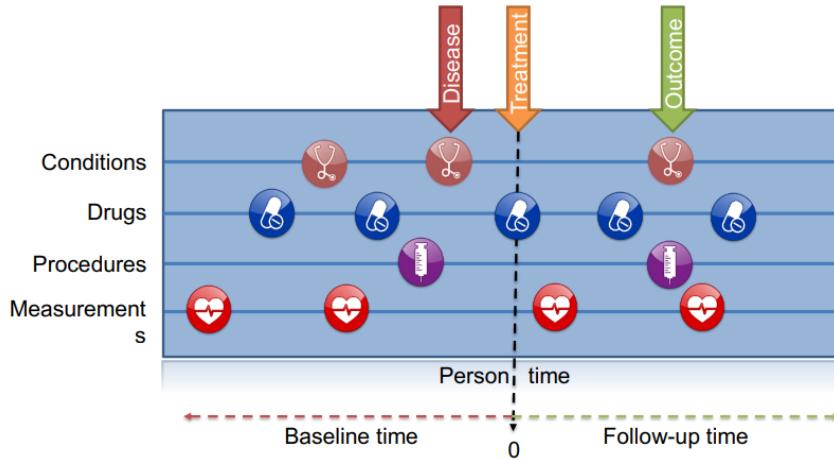


**Figura 5.7: The Journey from Data to Evidence.** Extraído del Libro de OHDSI [3]

La anterior Figura complementa a la Figura 5.4 con mayor información sobre los diferentes tipos de datos que se almacenan y los diferentes tipos de evidencia que se quiere generar. Frente a la disparidad de los datos y la finalidad con la que son recodigos, OHDSI presentará el Modelo de Datos Común (véase 5.4.1) y para la extracción de evidencia, la investigación metodológica mediante tres casos de uso fundamentales: la caracterización clínica de una cohorte, la estimación a nivel de población y la predicción a nivel de paciente (véase 5.4.3).

Los estudios que se llevan a cabo son estudios observacionales o fenotípicos, es decir, que simulan lo que sería un estudio clínico experimental sin intervención pero sobre datos ya almacenados de pacientes, en vez de realizar un seguimiento en vivo. De esta forma se logra obtener evidencia basada en datos. Cuando la evidencia se extrae sobre datos del mundo real (RWD), se denomina evidencia del mundo real (*Real World Evidence, RWE*).

Además, una característica fundamental de las investigaciones que se llevan acabo en OHDSI es que giran entorno al paciente, que es además el núcleo central del Modelo de Datos Común de OMOP. Por lo que un componente fundamental de cualquier investigación metodológica son las historias de los pacientes. **Para cada evento clínico que sucede se recoge una historia del paciente o *Patient Journey*.** Es importante no confundir este concepto con la Historia Clínica Electrónica de un paciente (HCE) que recoge una ficha con todos los eventos que han le sucedido a lo largo de su vida. Las investigaciones observacionales se diseñan para extraer información sobre la recopilación de todos los *patient journeys*.



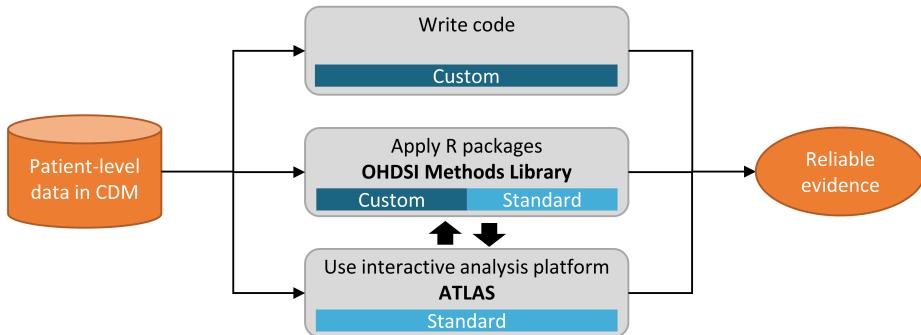
**Figura 5.8: The patient Journey.** Extraído de la página web oficial [3]

La historia del paciente, como se muestra en la Figura 5.8, es por tanto, una ventana temporal que recoge un evento clínico que le sucede a un paciente en un período de tiempo concreto. El evento se describe mediante tres períodos de tiempo: la enfermedad (rojo), el tratamiento (naranja) y el efecto (verde); y a partir de cuatro características esenciales: síntomas (*conditions*), medicamentos (*drugs*), procedimientos (*procedures*) y pruebas (*measurements*).

Los datos de pacientes se pueden agrupar en **cohortes** que comparten historias y características similares, al igual que a la hora de realizar un estudio clínico en vivo. Las diferentes prácticas para los análisis de cohortes darán lugar a los diferentes tipos de evidencia deseada (caracterización, estimación a nivel de población, predicción a nivel de paciente). **Por tanto, el componente central para generar evidencia en OHDSI será la cohorte.**

### Implementación del análisis

Para realizar un análisis, OHDSI distingue tres vías alternativas para generar la evidencia a partir de la base de datos estandarizada al OMOP CDM. Estas tres alternativas se muestran a continuación en la Figura 5.9, extraída del capítulo 8 del Libro de OHDSI.



**Figura 5.9:** Tres vías para la implementación de un análisis observacional. Extraído del Libro de OHDSI [3]

Cada vía se evalúa en cuanto a lo personalizada (*custom*) o estandarizada (*standard*) que es. A estas alturas se debe conocer que la vía más recomendada para implementar el análisis será la más estandarizada, es decir, la tercera vía.

La problemática que presentan la primera y la segunda vía consiste en ser en mayor o menor medida vías customizada, lo que genera problemas de interoperabilidad y reproducibilidad de los estudios. Si bien la primera vía consiste en la programación directa de código para realizar las consultas (no hay ningún tipo de estandarización, distintos lenguajes de programación, funciones personalizadas) al menos la segunda vía hace uso de librerías estándares en R que OHDSI ofrece (*OHDSI Methods Library*) pero, aunque se use el mismo lenguaje de programación y funciones, los scripts pueden ser tan distintos que aún dificulten la interoperabilidad.

Por tanto, la tercera vía se presenta como la alternativa óptima por ser la más estandarizada y es la que empleará el TFG en el estudio práctico. Esto es, usar la herramienta interactiva *low-code* de análisis de datos que ofrece OHDSI, denominada **ATLAS**, sin necesidad de programar directamente código.

## 5.4. Estándares

En términos de estandarización, OHDSI realiza una labor muy importante para paliar las dificultades de la investigación con datos de salud a causa de la heterogeneidad de los datos y estudios. Debido a la amplia colaboración internacional de la organización se reconoce la necesidad de estándares que permitan el intercambio de información sin pérdida entre los distintos sistemas de información de los miembros.

La estandarización de la investigación en OHDSI se realiza en dos planos: a nivel de datos y a nivel de investigación. La estandarización a nivel de datos se realiza a través del uso del Modelo de Datos Común 5.4.1 y el Vocabulario 5.4.2 y la estandarización a nivel de investigación mediante la investigación metodológica 5.4.3.

### 5.4.1. El Modelo de Datos Común

El Modelo de Datos Común o *Common Data Model* de OMOP es "un estándar de datos comunitario abierto, diseñado para estandarizar la estructura y el contenido de los datos de observación y permitir análisis eficientes que puedan producir evidencia confiable" [4], en definitiva, es un modelo semántico estándar para estructurar los datos de salud. La información más relevante y actualizada sobre el CDM se encuentra en su página de github [4] y en el capítulo 4 del Libro de OHDSI [3].

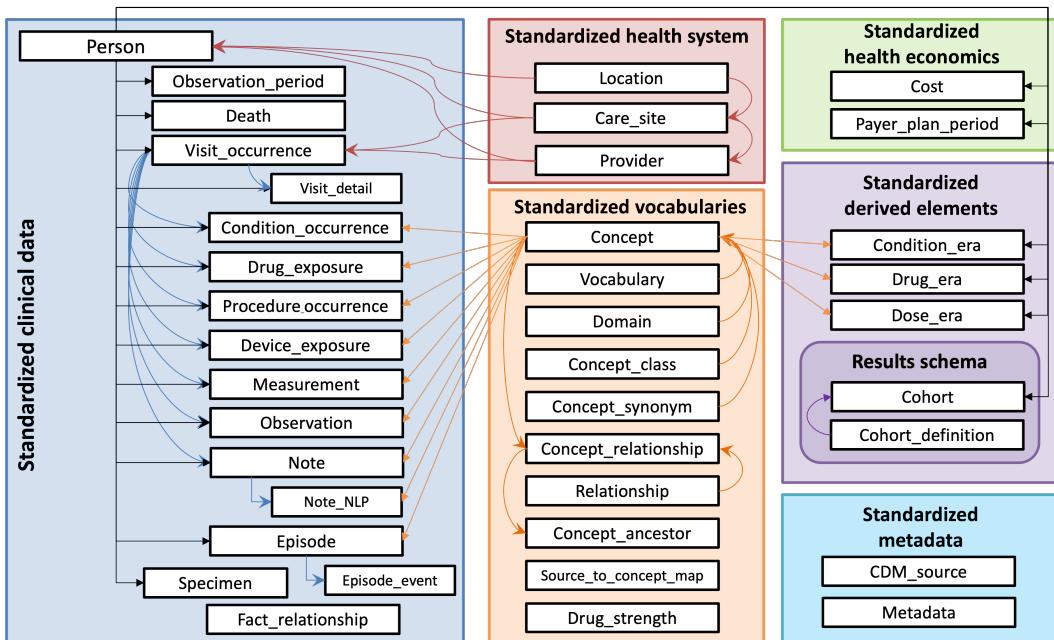
#### Características

El modelo de datos de OMOP presenta características importantísimas para hacer frente a las necesidades del panorama socio-sanitario actual presentado en 1.2. A continuación se presentan las características más relevantes del modelo (extraídas de la sección 4.1 del Libro de OHDSI [3]), según las necesidades identificadas previamente.

- **Estructura diseñada para la investigación.** El modelo presenta una estructura única y óptima para un propósito concreto: el de facilitar la realización de estudios observacionales. Por tanto reduce notoriamente los desafíos relativos a las diferentes estructuras y propósitos con los que se recogen los datos clínicos.
- **Modelo centrado en el paciente.** Es un modelo centrado en el paciente (alineado con la misma característica de la Sanidad 4.0). Estructuralmente esto significa que todos los eventos y tablas están relacionados con la tabla central del paciente, denominada *Person*.
- **Protección y privacidad.** El modelo limita el acceso a la información personal de los pacientes, evitando en la medida de lo posible el acceso a información personal sensible como nombres o fechas de nacimiento, para fomentar la protección y privacidad de los datos. Mayor información sobre las técnicas empleadas para ello se encuentran en el apartado *Privacidad del paciente y OMOP* de la página de github [4].
- **Reutilización de estándares.** Un aspecto importantísimo es que el modelo propone su propio estándar pero sin olvidar los estándares globalmente utilizados, de manera que integra y reutiliza los conceptos provenientes de estándares ya existentes (ej. SNOMED, LOINC...) referenciándolos en su propio Modelo de Datos Común. El conjunto de todos los estándares existentes adheridos al modelo de OMOP conforma el Vocabulario.
- **Neutralidad tecnológica.** El modelo no requiere una tecnología específica sino que puede estructurarse en cualquier base de datos relacional (ej. Oracle, SQL Server...), ajustándose a los requisitos tecnológicos necesarios de cada organización (identificado también como una dificultad en 1.2).

## Modelo de Datos Lógico

Actualmente el CDM ha lanzado ya su sexta versión, sin embargo, esta aún no está soportada por todas las herramientas de la comunidad, por lo que se sigue sugiriendo el uso del CDM v5.4, que es la última versión completamente funcional. A continuación, en la Figura 5.10 se presenta la estructura lógica de este modelo.



**Figura 5.10:** Estructura del CDM v5.4. Extraída de la página de github [4]

Quizás añadir explicación más detallada de tablas y/o relaciones más relevantes cuando empiece a manejar ATLAS

### 5.4.2. El Vocabulario

El Vocabulario es uno de los elementos centrales del Modelo de Datos Común de OMOP y una gran herramienta de estandarización e interoperabilidad entre sistemas. Como se comentaba en varias ocasiones, actualmente hay muchos estándares distintos en funcionamiento que establecen las terminologías de los eventos clínicos, como son FHIR, SNOMED CT, RxNorm u otros. El beneficio del Vocabulario de OMOP es que integra todos los vocabularios ya existentes en un único Vocabulario estándar, a través de la referencia entre conceptos estándar (pertenecientes a OMOP) y conceptos no estándar (pertenecientes a vocabularios alternativos).

El Vocabulario de OHDSI, por tanto, impera sobre un conjunto de vocabularios, respetando las diversas procedencias de cada término pero mapeándolos a un único vocabulario estándar. Como todas las herramientas de la comunidad, la información acerca de este está disponible online de forma pública en el capítulo 5 del Libro de

OHDSI [3] y en la página de github del CDM [4]. Por otra parte, existe un buscador online de términos en el Vocabulario de OMOP denominado ATHENA [46].

The screenshot shows the ATHENA search interface. At the top, there's a green header with the ATHENA logo and navigation links for SEARCH, DOWNLOAD, and user profile (Maria del Valle Ortiz). A search bar contains the keyword 'aspirin'. Below the header is a sidebar with filters for DOMAIN, CONCEPT, CLASS, VOCAB, and VALIDITY. The main area displays a table of search results for 'aspirin', with columns for ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The results show various entries like 'IKheis Local Municipality', 'IKheis Ward 1', and 'Section\* examination - approved doctor'. A footer at the bottom left includes a 'CLEAR FILTERS' button and the URL 'https://athena.ohdsi.org'.

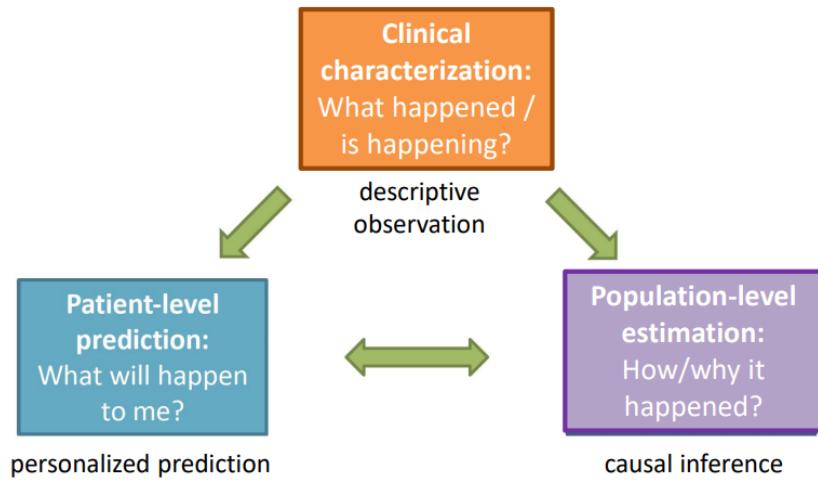
**Figura 5.11:** Captura de pantalla del menú principal de ATHENA

Actualmente hay más de nueve millones de términos registrados en el Vocabulario de OMOP, como se muestra en la Figura 5.11, y 155 vocabularios distintos coexisten juntos en el estándar de los cuales al menos 30 son vocabularios internos de OMOP.

Quizás añadir explicación más detallada de conceptos más relevantes cuando empiece a manejar ATLAS

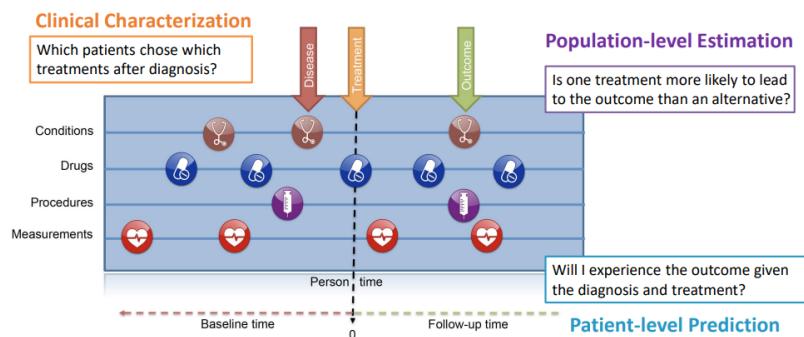
#### 5.4.3. Investigación metodológica

Con el fin de estandarizar y proveer un marco metodológico en el camino hacia la generación de evidencia, OHDSI define tres casos de usos que establecen los fines investigacionales sobre los datos del *Common Data Model*: (i) la caracterización, (ii) la estimación a nivel de población (iii) la predicción a nivel de paciente.



**Figura 5.12:** Esquema simplificado de los casos de uso para la investigación en OHDSI. Extraído del Symposium 2023, publicado en la web oficial [2]

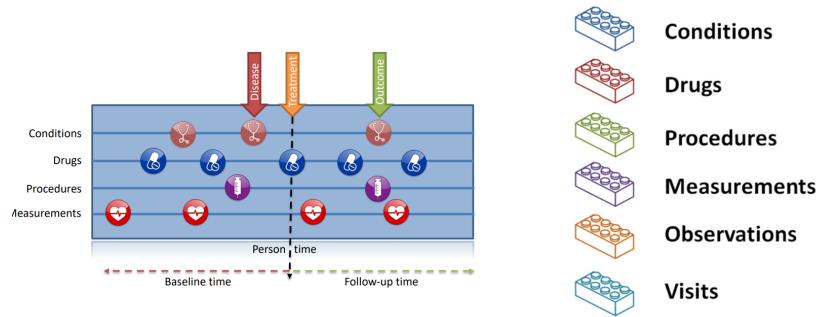
Estos tres casos de uso se presentan generalmente en los *Symposium* y *Workshops*, para dar a conocer a la comunidad el esquema propio de investigación de OHDSI. Además el Libro de OHDSI [3] los presenta de forma general en el capítulo 7 y de forma específica para cada caso de uso en los capítulos 11, 12 y 13, respectivamente. A continuación se muestra otro esquema más complejo también presentado en el Symposium 2023 en el que se encuadra cada caso de uso en la ventana temporal que encuadra un estudio observacional.



**Figura 5.13:** Esquema de los casos de uso encuadrado en la historia del paciente. Extraído del Symposium 2023 publicado en la web oficial [2]

## Cohortes

Añadir explicación más detallada con conceptos relevantes cuando empiece a manejar ATLAS



**Figura 5.14:** Building blocks para la definición de una cohorte. Extraído del Tutorial 2022 publicado en la web oficial [2]

## Caracterización

Añadir explicación más detallada con conceptos relevantes cuando empiece a manejar ATLAS

La caracterización busca la caracterización a nivel estadístico de un cohorte o una base de datos. Es una mera descripción estadística de los datos, sin realizar inferencias, predicciones o análisis más complejos, simplemente observando la base de datos.

La caracterización consiste en caracterizar una población a través de métricas estadísticas descriptivas con el objetivo de formular hipótesis sobre los determinantes de la salud y la enfermedad.

## Estimación a nivel de población

Añadir explicación más detallada con conceptos relevantes cuando empiece a manejar ATLAS

La estimación a nivel de población busca realizar inferencias causales sobre los efectos de las intervenciones sanitarias en la población. Se pretende entender los efectos causales para comprender las consecuencias de las acciones

## Predictión a nivel de paciente

Añadir explicación más detallada con conceptos relevantes cuando empiece a manejar ATLAS

La predicción a nivel de paciente busca, en base a los datos obtenidos de los conjuntos de pacientes en la base de datos, realizar predicciones concretas para un individuo concreto.

## 5.5. Herramientas

OHDSI proporciona un conjunto de herramientas para facilitar la realización de los estudios e investigaciones a raíz de los datos clínicos y fomentar la interoperabilidad entre estos, aportando un estándar de herramientas.

Las herramientas que proporciona la organización están disponibles públicamente online y de forma gratuita y son desarrolladas por los propios miembros de la comunidad. Entre todas las herramientas, para la realización de este Trabajo Fin de Grado se destaca la herramienta de análisis de datos clínicos ATLAS [5.5.1](#), aunque también existen otras herramientas importantes de forma indirecta [5.5.2](#).

### 5.5.1. ATLAS

ATLAS es la herramienta de OHDSI por excelencia porque es la que estandariza el análisis observacional una vez que la base de datos está convertida al modelo OMOP. La documentación oficial sobre ATLAS se encuentra en el capítulo 8 del Libro de OHDSI y en su repositorio de github [\[5\]](#). Además, aparte de la documentación oficial, hay montones de información esparcidas por la red sobre ATLAS, en publicaciones científicas, foros de OHDSI, videotutoriales en youtube y un largo etcétera.



**Figura 5.15:** Logo de ATLAS. Extraída del repositorio de github [\[5\]](#)

Un importante promotor del uso de ATLAS es la red europea de datos EHDEN [\[47\]](#) (véase [1.3](#)). En esta línea, también la plataforma EHDEN Academy también ofrece cursos gratuitos sobre el uso de ATLAS y otras herramientas OHDSI.

#### Características y beneficios de su uso

El uso de ATLAS es beneficioso para la comunidad científica debido principalmente a su naturaleza *open-source, low-code* y la reproducibilidad que ofrece para los estudios:

- I. **Open source.** ATLAS se presenta como una herramienta disponible públicamente online, configurable gracias a su característica de código

abierto, que expone toda su información y el propio código que la compone en los repositorios de github de la organización y, por si fuera poco, cuenta con el apoyo de un equipo de desarrolladores pendiente en los foros e *issues* que se reportan vía github para solucionar las dudas que tengan los implementadores.

- II. **Low-code.** Por otro lado, no requiere de conocimientos expertos de programación, puesto que es *low-code*. La herramienta se implementa sobre la Biblioteca de Métodos de OHDSI, con soporte para análisis en R, pero no requiere programación directa sino que ofrece una interfaz gráfica e intuitiva para el analista de datos. Además, el código que subyace al análisis es fácilmente exportable, siempre estructurado según el mismo estándar, favoreciendo la interoperabilidad del mismo.



**Figura 5.16:** Biblioteca de Métodos OHDSI que utiliza ATLAS. Extraída del Libro de OHDSI [3]

Todo ello no solo facilita la tarea del analista de datos sino que además favorece la interoperabilidad entre los estudios, puesto que todos los estudios que utilizan ATLAS implementan (en una capa inferior) los mismos métodos, el mismo lenguaje de programación y la misma estructura de análisis (véase 5.3).

- III. **Reciclabilidad.** Por último, otro beneficio es que gracias a estas características ATLAS permite diseñar estructuras para el estudio de los datos que puedan utilizarse en diferentes bases de datos distintas. Volviendo al ejemplo de la plancha en 5.3, esto quiere decir que una misma plancha (o estudio) puede conectarse a cualquier enchufe de cualquier región (a cualquier base de datos). ATLAS está intrínsecamente configurada para diseñar análisis reproducibles, por lo que los elementos que se configuran durante un análisis de datos (grupos de cohortes, estimadores, predictores, grupos de

conceptos...) se pueden exportar fácilmente a modo de estructura general e implementarse sobre otro estudio que, aunque posea datos distintos ejecute ATLAS. Por tanto las estructuras más eficientes que se utilicen en un análisis remoto, pueden compartirse en la red de la comunidad y ser utilizados en cualquier nodo y cualquier estudio, favoreciendo la reciclabilidad, reproducibilidad e interoperabilidad del estudio.

### Aspectos técnicos

En cuanto a los aspectos técnicos, ATLAS se despliega como una herramienta basada en web, normalmente alojada en un servidor Apache, combinada con la WebAPI de OHDSI. Generalmente se recomienda su despliegue en Google Chrome. Además la herramienta puede implementarse de forma pública a través de internet o tras el firewall de la red privada de una organización, según las necesidades de la entidad que lo implementa.

Sin embargo, es importante recalcar que tanto ATLAS como la mayoría de las herramientas de OHDSI no consiste en un archivo ejecutable aislado sino en una aplicación contenida y dependiente de un ecosistema completo basado en web. La dependencia principal y red que sostiene a ATLAS es la **WebAPI**.

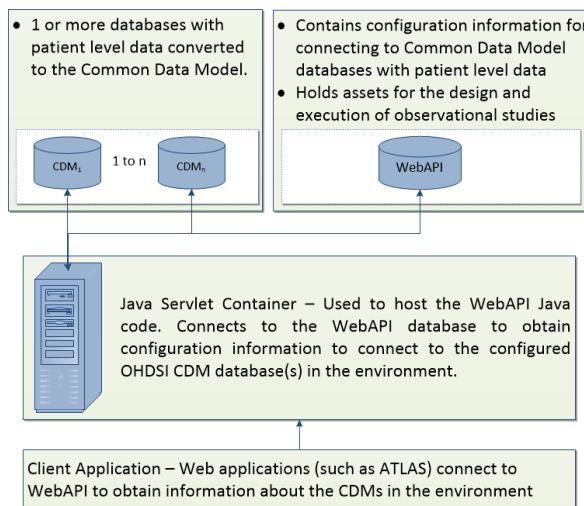
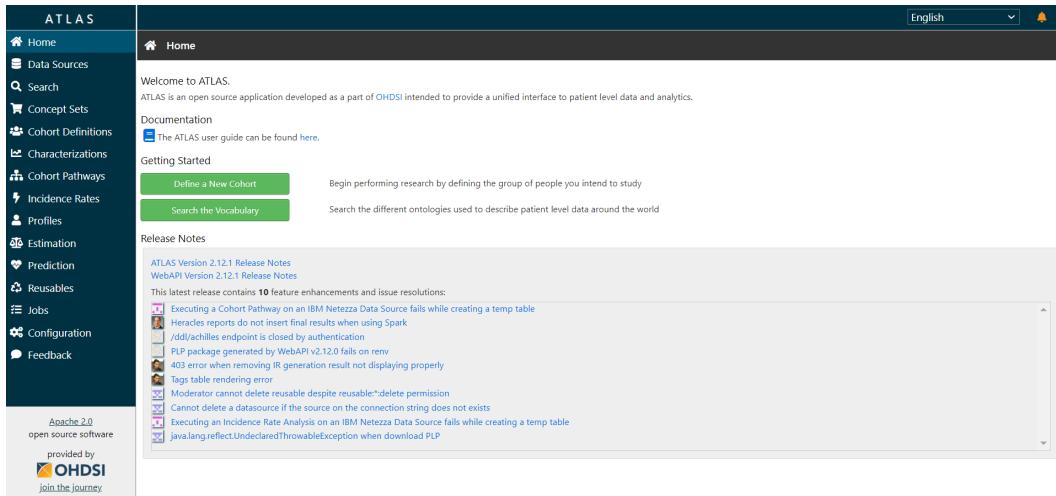


Figura 5.17: Estructura de la WebAPI. Extraída de la wiki de github [6]

Tal y como se muestra en la Figura 5.17, la Web API es la aplicación que proporciona los servicios RESTful para que la herramienta pueda interactuar con las bases de datos [6]. Por tanto su relación con ATLAS es estrictamente necesaria. ATLAS no es una herramienta aislada sino un eslabón del ecosistema OHDSI.

Por otra parte, la herramienta en sí se muestra a través de una interfaz gráfica, que proporciona un estrecho menú lateral con 15 herramientas para el análisis de datos. La interfaz de la herramienta seleccionada se muestra en el lado derecho, como se muestra en la Figura 5.18.



**Figura 5.18:** Captura de pantalla del menú principal de ATLAS demo

Recientemente, en diciembre de 2023, ATLAS lanzó su versión 2.14.1 que está en correcto funcionamiento y es la que se utiliza en el desarrollo del Trabajo Fin de Grado. Más información sobre los aspectos técnicos de la herramienta se encuentran en el repositorio de github [5].

## Estrategias de Implementación

La implementación de ATLAS en una organización puede ser una tarea complicada por su dependencia con la WebAPI, la Biblioteca de Métodos y otras dependencias al ecosistema OHDSI.

No obstante, la organización ha desarrollado varias iniciativas que facilitan su implementación y accesibilidad, para no crear obstáculos en la promoción del uso de la herramienta. Estas iniciativas se describen a continuación.

- ATLAS demo** [48]. En primer lugar, esta es una herramienta muy fácilmente accesible que proporciona la comunidad científica para tomar un primer contacto con la herramienta. En este caso, la herramienta es accesible a través del navegador web, públicamente a través de Internet. Cualquier usuario de internet tiene acceso a la herramienta demo. Se le denomina demo porque se sobreentiende que su uso es principalmente educativo o formativo, aunque verdaderamente ofrece todas las capacidades de la herramienta y los análisis que con ella se realizan, podrían reutilizarse en estudios más complejos o de organizaciones privadas.
- ATLAS Docker**. Por otro lado, también muy fácilmente implementable se presenta **Broadsea** [49], que consiste en la virtualización del ecosistema OHDSI en un multicontenedor Docker. Gracias a la facilidad del uso de las tecnologías Docker, esta forma de implementar el ecosistema es bastante sencilla, permitiendo además añadir nuevas configuraciones más complejas (si fuese necesario) añadiendo o eliminando contenedores. Para realizar la parte práctica de este trabajo se emplea la tecnología Docker de Broadsea

para implementar ATLAS. A la herramienta ATLAS desplegada con Broadsea, frecuentemente se le denominará a lo largo del documento *ATLAS Broadsea*. El TFG presenta un documento anexo bastante complejo enteramente dedicado a la instalación, despliegue y configuración del entorno Broadsea (véase anexo A). Además, la arquitectura de Broadsea también se presenta en 7.

- c. **ATLAS Amazon Web Services.** Otra alternativa que propone la organziación, en colaboración con Amazon, es la virtualización del ecosistema en el entorno de computación en la nube de Amazon Web Services (AWS). Para ello se ofrecen los entornos *OHDSI-in-a-Box* [50] y *OHDSIonAWS* [51]. OHDSI-in-a-Box se crea específicamente como un entorno de aprendizaje y se utiliza en la mayoría de los tutoriales proporcionados por la comunidad OHDSI mientras que OHDSIonAWS es una arquitectura de referencia para entornos OHDSI de clase empresarial, multiusuario, escalables. Por las restricciones intrínsecas al uso de AWS, estas alternativas han sido rechazadas para ser empleadas en el TFG.
- d. **ATLAS Azure.** Por último, *OHDSI on AZURE* [52] es otra alternativa de virtualización pero a través de la plataforma Microsoft de Azure. No obstante, esta alternativa es la menos común.

## Herramientas embebidas

Si bien las herramientas del ecosistema de OHDSI no son totalmente aisladas, ATLAS presenta en su propia interfaz acceso a dos de estas herramientas de forma íntegra, para facilitar la eficiencia y rapidez en el análisis. Estas herramientas son las siguientes:

- **ACHILLES** [53]. Esta herramienta, de las siglas *Automated Characterization of Health Information at Large-Scale Longitudinal Evidence Systems*, en español Caracterización automatizada de la información sanitaria en sistemas de evidencia longitudinal a gran escala, sirve para caracterizar y/o obtener un reporte estadístico de la base de datos estandarizada que se va a utilizar para el estudio. Intrínsicamente es una librería de R que se implementa como una opción del menú lateral *Data Sources* de ATLAS.
- **ATHENA** [54]. Esta herramienta sirve para realizar búsquedas dinámicas en el Vocabulario de OMOP (véase 5.4.2). Está implementada en ATLAS en la opción *Search* del menú lateral. Además, se puede acceder a ella online de forma externa a través de su propia página web [46].

### 5.5.2. Otras herramientas

El ecosistema de OHDSI presenta gran cantidad de herramientas adicionales. A continuación se presentan otras herramientas que aunque no se utilizan directamente, son importantes para realizar un análisis de datos completo.

- **HADES** [55]. HADES, del inglés *Health Analytics Data-To-Evidence Suite* y en español Suite de análisis sanitario de datos a evidencia, es el nombre con el que se denomina a la herramienta que implementa el paquete R con la Biblioteca de Métodos de OHDSI (ver Figura 5.16). Se puede instalar como un entorno independiente mediante Java y Rtools para implementar análisis mediante código estandarizado (véase Figura 5.9). No se utiliza en el TFG más allá de la implementación subyacente de las bibliotecas en ATLAS.
- **Rabbit tools y Usagi** [56]. Estas herramientas en conjunto llevan a cabo el proceso de ETL, para omopizar las bases de datos al Modelo de Datos Común de OMOP. Las herramientas son tres: White-Rabbit, Rabbit-In-Hat y Usagi. No se utiliza directamente en el TFG porque el dataset utilizado para el análisis ya estaba previamente omopizado.
- **Data Quality Dashboard** [57]. Esta herramienta, en español Panel de control de calidad de los datos, pertenece a un paquete de HADES aunque implementado como una interfaz gráfica aparte para facilitar su acceso online. Tal y como su nombre indica sirve para automatizar la tarea de comprobación de la calidad de los datos, un paso previo fundamental antes de realizar un análisis de datos. Tampoco se utiliza directamente para el TFG porque este estudio se llevó a cabo durante la omopización del dataset.

## 5.6. Conclusiones

En este capítulo se concluye la relevancia de OHDSI en el panorama sanitario y tecnológico actual, capaz de subsanar las necesidades del mismo a través de su misión, estándares y herramientas (5.2, 5.4, 5.5).

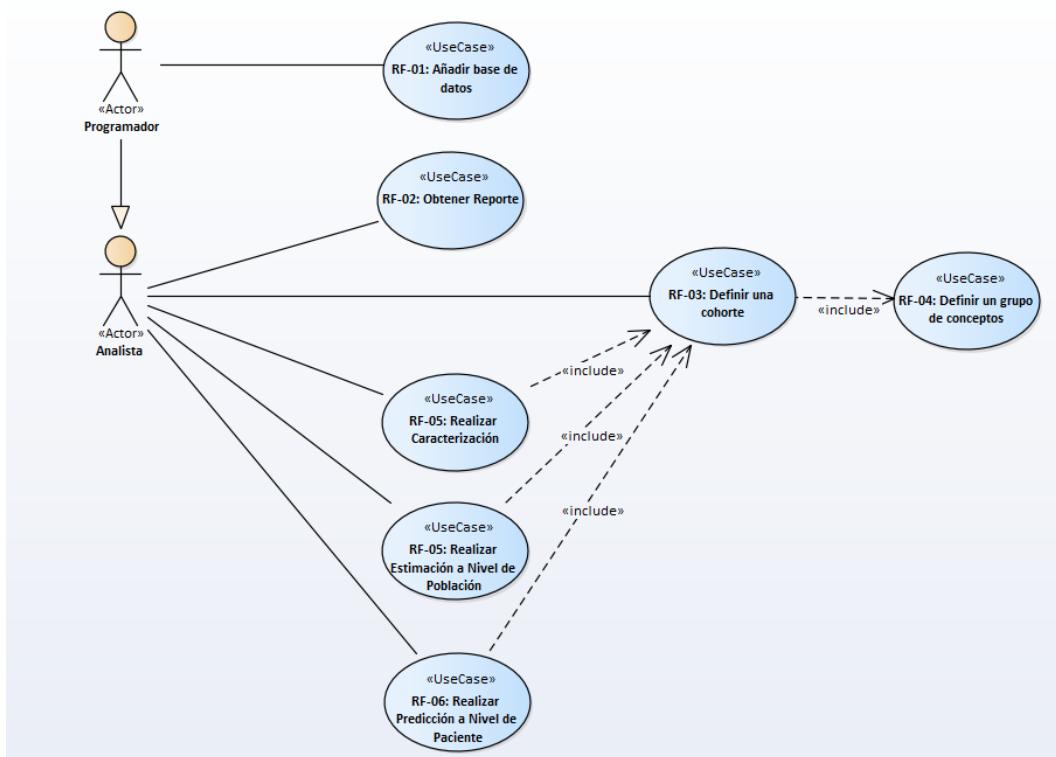
Principalmente se realiza una labor importantísima de estandarización, tanto a nivel de datos mediante el Modelo de Datos Común 5.4.1 y el Vocabulario 5.4.2 como a nivel de estudios 5.4.3 mediante la Investigación metodológica y sobre todo gracias al auge del uso de su herramienta estrella, ATLAS 5.5.1, que se emplea para el estudio práctico del TFG.

# 6. Documento de Requisitos

## 6.1. Introducción

## 6.2. Requisitos Funcionales

### 6.2.1. Diagrama de casos de uso



**Figura 6.1:** Boceto de diagramas de casos de uso

### 6.2.2. Casos de uso

- RF01: Cargar datasets
- RF02: Obtener un reporte del data set (Data source)
- RF03 : Definir conjuntos de conceptos del Vocabulario (Concept set)
- RF04: Configurar la muestra de trabajo (cohort definition)
- RF05: Caracterizar el cohort (characterization - primer gran bloque de metodología de OHDSI)

- RF06: Definir una estimacion a nivel de poblacion (estimation - segundo gran bloque de metodologia de OHDSI)
- RF07: Hacer una prediccion a nivel de paciente (prediction - tercer gran bloque de metodologia de OHDSI)

## **6.3. Requisitos no Funcionales**

## **6.4. Conclusiones**

En este capítulo concluimos que...

# 7. Entorno de Trabajo

---

Este capítulo presenta el entorno de trabajo: [7.2 la arquitectura tecnológica](#) y [7.3 el entorno tecnológico](#).

## 7.1. Introducción

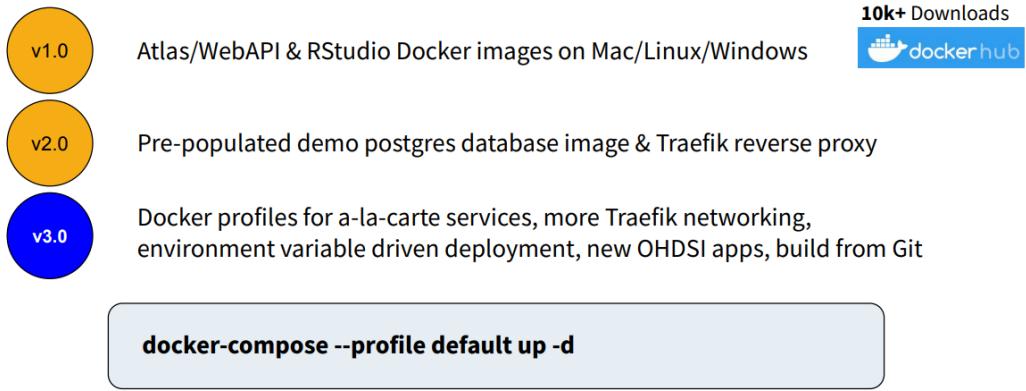
La implementación del ecosistema de herramientas OHDSI y ATLAS puede ser una ardúa tarea. En el contexto de desarrollo del Trabajo Fin de Grado junto a las prácticas en empresa en el Hospital Virgen del Rocío, la dificultad de la tarea se ve exponencialmente aumentada debido a los grandes protocolos de seguridad y privacidad de la administración pública. Por ello, se ha seleccionado el despliegue de las herramientas OHDSI a través del subsistema Docker de Broadsea, que presenta una vía sencilla para realizar esta labor.

Broadsea es un proyecto basado en Docker que permite desplegar todo el entorno de herramientas, configuraciones y dependencias OHDSI de la manera más sencilla hasta el momento. De hecho, la misma organización la presenta, textualmente, como "*la forma más sencilla de instalar (y actualizar) las herramientas OHDSI*" [58].



**Figura 7.1:** Esquema sencillo de Broadsea. Extraída de [\[7\]](#).

Aunque comenzó en su primera versión, como un simple contenedor que albergaba imágenes de la WebAPI de ATLAS y RStudio [7] ha evolucionado hasta la tercera versión en la que Broadsea alberga la mayoría de herramientas OHDSI, creando un entorno virtual de desarrollo muy completo (véase Figura [7.4](#)).



**Figura 7.2:** Historial de versiones de Broadsea. Extraída de [7].

Como se muestra en la Figura 7.2, la tercera versión de Broadsea incluye perfiles docker que facilitan la configuración del servicio, redes internas de conexión, despliegue de variables internas, más aplicaciones OHDSI y la construcción de contenedores desde Git, entre otros.

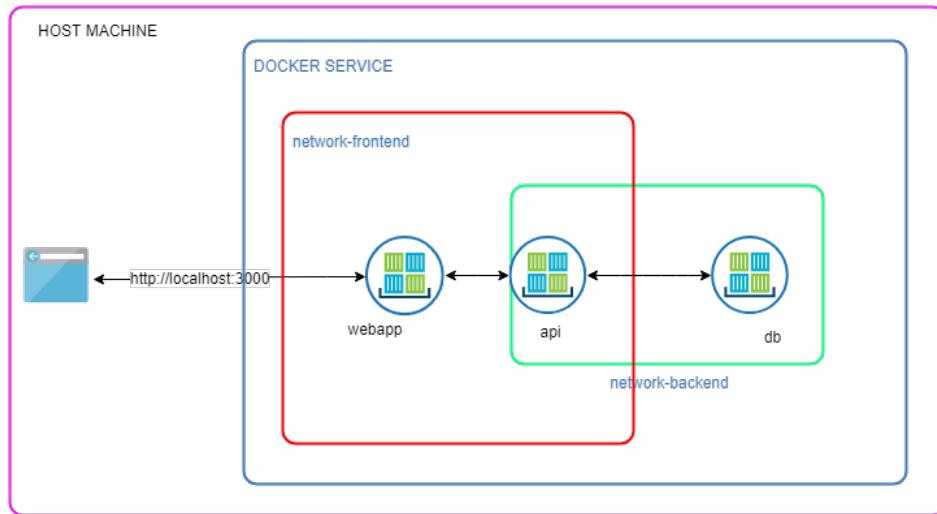
A continuación se presenta la arquitectura teórica de Broadsea y el entorno de herramientas tecnológicas necesarias para su correcta implementación.

## 7.2. Arquitectura tecnológica

La arquitectura en términos tecnológicos del sistema es compleja, por ello se describe en dos subsecciones: 7.2.1 Arquitectura teórica-generalizada del sistema y 7.2.2 Arquitectura específica de Broadsea.

### 7.2.1. Arquitectura del sistema

El sistema se implementa mediante virtualización con Docker y una arquitectura en tres niveles o *three-tier*, donde se diferencian al cliente, frontend y backend. Esta arquitectura se describirá de forma general utilizando el esquema de la Figura 7.3, extraído de internet.



**Figura 7.3:** Esquema de arquitectura *three-tier* en Docker. Extraída de [8].

En primer lugar, la virtualización obliga a diferenciar entre una maquina local o anfitriona (*host machine*, en rosa) y una maquina virtual que provee el servicio docker (*docker service*, en azul).

1. **La máquina local.** La máquina local es la propia máquina del usuario. Se le denomina anfitriona porque aloja en su interior a la máquina virtual. La máquina local cede un servidor y un puerto a la máquina virtual para que el usuario final pueda acceder al sistema a través de la dirección del servidor en que se aloja, típicamente accediendo mediante un navegador web. El acceso mediante el navegador web es lo que se denomina la capa cliente, pues es la interfaz que permite al usuario acceder al sistema.
2. **La máquina virtual.** La máquina virtual es el sistema virtualizado en Docker. Es el sistema que contiene toda la lógica de la aplicación y los datos empaquetado en un multicontenedor Docker, en este caso el multicontenedor es el propio sistema Broadsea. Está compuesto por tres nodos la *webapp*, la *api* y la *db* que conforman las dos capas restantes de la arquitectura: el frontend y el backend.

Por tanto, a nivel de arquitectura del sistema en sí, se encuentra la capa cliente (en el *host machine*, en rosa), el frontend (*network-frontend*, en rojo) y el backend (*network-backend*, en verde).

1. **El cliente.** El cliente está alojado en la máquina anfitriona y proporciona el acceso a los servicios virtualizados del sistema a través de la conexión internet con el servidor docker.

En el caso de Broadsea el navegador deberá ser Google Chrome y la dirección por defecto será <http://127.0.0.1:5432>.

2. **El frontend.** El frontend está alojado en la máquina virtual, es el servicio que guarda la lógica de la aplicación que se muestra al usuario. Se compone de la *webapp*, que contiene la aplicación como tal, y la *api*. que es la red que permite

establecer interconexiones entre la aplicación lógica y la base de datos; entre el frontend y el backend.

En el caso de Broadsea la webapp y la api se combinan en el componente de la WebAPI, que permite el acceso a la aplicación de ATLAS y maneja las conexión con las bases de datos del backend.

3. **El backend.** El backend está alojado en la máquina virtual, es el servicio que aloja la base de datos sobre la que se sostiene la aplicación. Se compone de la *api* y la *db*. De igual forma que en el frontend, la api es la red que permite la interconexión entre los componentes del sistema, en este caso con la base de datos, que puede ser una o varias.

En el caso de Broadsea, las bases de datos deberán estar estandarizadas a OMOP y podrán encontrarse en el propio servidor Docker, como es el caso de Eunomia, o en servidores externos. No obstante, la relación entre cualquier base de datos y ATLAS se realiza a través de la WebAPI.

### 7.2.2. Arquitectura de Broadsea

A continuación se presenta la arquitectura específica de Broadsea. Broadsea v3.0 es un sistema muy complejo, contenido en un multicontenedor Docker que alberga el ecosistema completo de herramientas OHDSI y sus interconexiones en distintos contenedores. Además, se definen distintos perfiles (*profiles*) para facilitar la instalación de los distintos contenedores. Por ello se le denomina *a-la-carte*. La figura 7.4 muestra todos los contenedores de Broadsea.

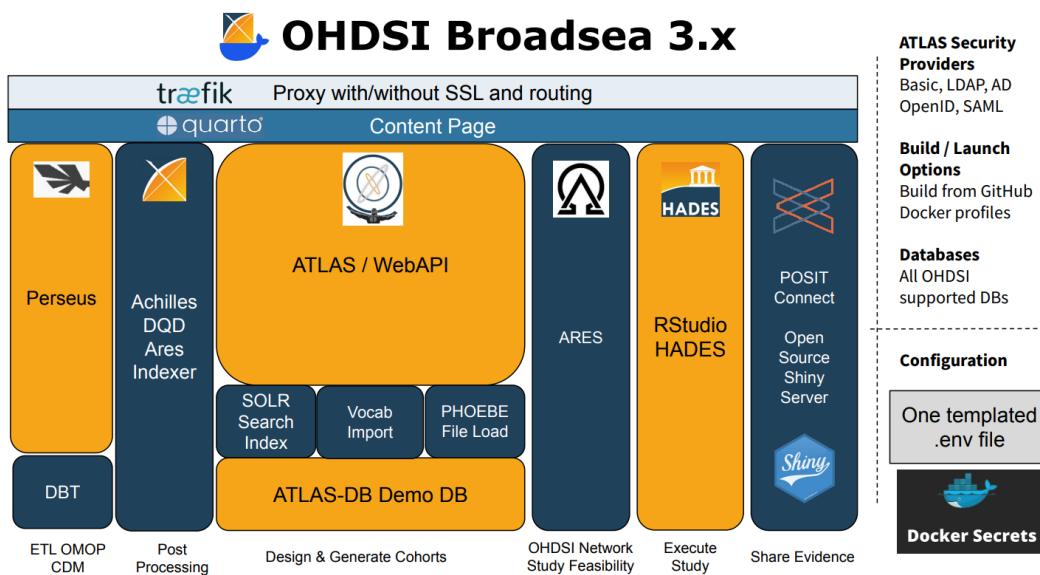


Figura 7.4: Vista general de todos los componentes de Broadsea. Extraída de [7].

El despliegue por defecto de Broadsea genera en el frontend una interfaz de usuario con acceso a tres aplicaciones: ATLAS, HADES y ARES. Para acceder a esta interfaz

de usuario basta con buscar en el navegador el servidor y puerto donde se aloja broadsea, que tipicamente será la maquina local y el puerto 5354, correspondiente a Postgre.

1. **ATLAS.** ATLAS Broadsea despliega todas las funcionalidades de la herramienta de forma local. ATLAS se sostiene sobre la WebAPI y cuenta con la base de datos de Eunomia.
  - a) **WebAPI.** La WebAPI se despliega como un contenedor docker y como un volumen de datos. Además, también se construirá un esquema en la base de datos del servidor Postgre que aloja al contenedor, denominado webapi. A través de la modificación de este esquema se podrán agregar o eliminar las diferentes fuentes de datos a la herramienta.
  - b) **BD.** Para facilitar el correcto funcionamiento de ATLAS se implementa una base de datos demo que es Eunomia. Esta base de datos cuenta con un pequeño registro de datos normalizados a OMOP y también crea varios esquemas en la base de datos del servidor Postgre que permiten su configuración, o la realización de consultas directamente desde el administrador de la base de datos.
2. **HADES.** HADES Broadsea despliega todas las funcionalidades de la herramienta de forma local. Se sostiene sobre una virtualización del IDE de RStudio que tiene preinstalada y preconfiguradas todas las librerías de la Librería de Métodos. Su uso no es relevante en el TFG.
3. **ARES.** ARES Broadsea despliega todas las funcionalidades de la herramienta de forma local. Su uso tampoco es relevante en el TFG.

## 7.3. Entorno tecnológico

El entorno tecnológico que envuelve al sistema está compuesto por cuatro herramientas fundamentales: Docker, PostgreSQL, Google Chrome y Github.

### 7.3.1. Google Chrome

Google Chrome es el navegador web de Google que permite el acceso a internet y la búsqueda en la web a través de una interfaz amigable e intuitiva [59].

El uso de Chrome para el acceso a Broadsea es muy recomendado por los propios desarrolladores, por lo que es una herramienta de especial relevancia.

### 7.3.2. Docker

Docker es una plataforma abierta para desarrollar, enviar y ejecutar aplicaciones. Docker le permite separar sus aplicaciones de su infraestructura para que pueda

entregar software rápidamente. Con Docker, puede administrar su infraestructura de la misma manera que administra sus aplicaciones [60].

De esta forma, Docker permite empaquetar y ejecutar aplicaciones en contenedores, entornos poco aislados pero seguros. Esto posibilita la ejecución de múltiples contenedores simultáneamente en un mismo host, sin depender de lo instalado en él. Los contenedores son ligeros y contienen todo lo necesario para la aplicación, facilitando su compartición y asegurando consistencia entre usuarios.

El uso de Docker en el desarrollo del TFG es evidente, es la herramienta que despliega Broadsea y, por consiguiente, ATLAS (véase ??). El proceso concreto de instalación, despliegue y configuración de Docker así como la explicación detallada de su estructura y archivos más importantes se presenta en el anexo A.

### 7.3.3. PostgreSQL

PostgreSQL es un potente sistema de base de datos relacional de objetos de código abierto que utiliza y amplía el lenguaje SQL combinado con muchas funciones que almacenan y escalan de forma segura las cargas de trabajo de datos más complicadas [61].

El uso de postgres es fundamental para la implementación correcta de Broadsea, puesto que la WebAPI se implementa sobre un sistema PostgreSQL. Las bases de datos externas que se interactúan con la WebAPI pueden estar en otros lenguajes relacionales, pero el sistema de Broadsea intrínsecamente solo se sostiene sobre Postgre.

De nuevo, el proceso concreto de instalación, despliegue y configuración de la base de datos Postgre así como la explicación detallada de su estructura y archivos más importantes se presenta en el anexo A.

### 7.3.4. Github

GitHub es una plataforma para desarrolladores que les permite crear, almacenar, gestionar y compartir su código. Utiliza el software Git, proporcionando control de versiones distribuido, además de control de acceso, seguimiento de errores, solicitudes de funciones de software, gestión de tareas, integración continua y wikis para cada proyecto [62].

El uso de Github es muy recomendado debido a que la mayor parte de la información sobre OHDSI y sus herramientas se encuentran en internet disponibles en repositorios de Github (véase 5.2).

Además, siguiendo esta iniciativa de OHDSI, para desarrollar este Trabajo Fin de Grado se ha creado un repositorio de Github específico [27] que contiene toda la documentación relevante a su desarrollo (archivos latex, pdf...) y archivos de variables de entorno o scripts utilizados durante la configuración del entorno del sistema o la realización del análisis de datos.

## 7.4. Conclusiones

En este capítulo se concluye que la arquitectura tecnológica del sistema es bastante compleja (véase [7.2](#)), puesto que involucra una virtualización del ecosistema OHDSI a través de Docker, denominado Broadsea. No obstante, la implementación del sistema en Docker facilita bastante la tarea de configurar el ecosistema completo, gracias al empaquetamiento de las funcionalidades en contenedores accesibles *a-la-carte*. Por último, es importante conocer el entorno tecnológico fundamental para desplegar correctamente el sistema, compuesto por Chrome, Docker, PostgreSQL y Github (véase [7.3](#)).

# **8. Caso práctico**

---

## **8.1. Introducción**

Se va a reproducir un estudio oncológico realizado por los investigadores del HUVR pero utilizando herramienta ATLAS

## **8.2. Estudio del HUVR**

Presentación del estudio oncológico realizado por los investigadores del HUVR (no utiliza ATLAS)

- Dataset utilizado
- Metodología si fuese relevante
- Objetivos del estudio
- Resultados

## **8.3. Reproducción del estudio**

### **8.3.1. Datos**

Se han obtenido el dataset que se utilizó en ese estudio.

- Los datos se convirtieron a OMOP (TFG Paco)
- Se analizó la calidad de los datos eficiente (TFG Paco)

### **8.3.2. Metodología**

#### **Reporte del dataset**

#### **Definición de la cohorte**

Necesario para utilizar ATLAS

Cohorte = Personas con cancer de pulmon

### **Caracterización de la cohorte**

Para conocer los pacientes que tenemos en el cohorte

### **Estimación a nivel de población**

Hacer un estudio sobre los efectos adversos que sufrirá la población del cohorte

### **Predicción a nivel de Paciente**

Para un paciente

#### **8.3.3. Resultados**

### **8.4. Comparación de resultados**

Comparación de los resultados obtenidos en el estudio del HUVR y el estudio realizado con ATLAS

### **8.5. Conclusiones**

En este capítulo concluimos que...

# **9. Resultados**

---

## **9.1. Resultados**

## **9.2. Trazabilidad de objetivos**

Trazabilidad de objetivos con resultados

## **9.3. Lecciones aprendidas**

- Comprensión de la importancia de la estandarización (estandar OHDSI) en la interoperabilidad de los sistemas clínicos.
- Implementación de un entorno virtual en el PC (Entorno y webAPI de OHDSI en MV DOCKER)
- Aprendizaje de uso de la herramienta ATLAS

...

## **10. Conclusiones**

---

# Bibliografía

---

- [1] Scrum. Scrum home page. URL <https://www.scrum.org/>.
- [2] Observational Health Data Sciences and Informatics. Ohdsi.org. <https://www.ohdsi.org/>,.
- [3] Observational Health Data Sciences and Informatics. The book of ohdsi. <https://ohdsi.github.io/TheBookOfOhdsi.html>, January 11 2021.
- [4] Observational Health Data Sciences and Informatics (OHDSI). Common data model, 2023. URL <https://ohdsi.github.io/CommonDataModel/index.html>.
- [5] OHDSI github. Atlas,. URL <https://github.com/OHDSI/Atlas>.
- [6] OHDSI github. Ohdsi webapi wiki, 2023. URL <https://github.com/OHDSI/WebAPI/wiki>.
- [7] A. Londhe. Slides from the 2023 ohdsi global symposium, 2023. URL <https://www.ohdsi.org/wp-content/uploads/2023/10/419-Londhe-Slides.pdf>.
- [8] MUNDANECODE. Three tier architecture in docker, 2024. URL <https://mundanecode.com/posts/three-tier-architecture-in-docker/>.
- [9] Heiner Lasi, Peter Fettke, Hans-Georg Kemper, Thomas Feld, and Michael Hoffmann. Industry 4.0: Towards future industrial opportunities and challenges. *Business & information systems engineering*, 6:239–242, 2014.
- [10] Chiehfeng Chen, El-Wui Loh, Ken N Kuo, and Ka-Wai Tam. The times they are a-changin’–healthcare 4.0 is coming! *Journal of medical systems*, 44:1–4, 2020.
- [11] Guilherme Luz Tortorella, Flávio Sanson Fogliatto, Alejandro Mac Cawley Vergara, Roberto Vassolo, and Rapinder Sawhney. Healthcare 4.0: trends, challenges and research directions. *Production Planning & Control*, 31(15):1245–1260, 2020.
- [12] Guilherme Luz Tortorella, Tarcísio Abreu Saurin, Flavio S Fogliatto, Valentina M Rosa, Leandro M Tonetto, and Farah Magrabi. Impacts of healthcare 4.0 digital technologies on the resilience of hospitals. *Technological Forecasting and Social Change*, 166:120666, 2021.
- [13] Susana Rubio Martín and Sonia Rubio Martín. ehealth y el impacto de la cuarta revolución industrial en salud, el valor del cuidado. *Enfermería en cardiología: revista científica e informativa de la Asociación Española de Enfermería en Cardiología*, (82):5–9, 2021.
- [14] Angelina Kouroubali and Dimitrios G Katehakis. The new european interoperability framework as a facilitator of digital transformation for citizen empowerment. *Journal of biomedical informatics*, 94:103166, 2019.

- [15] Rocío B Ruiz and Juan D Velásquez. Inteligencia artificial al servicio de la salud del futuro. *Revista Médica Clínica Las Condes*, 34(1):84–91, 2023.
- [16] Christina Ntafi, Stergiani Spyrou, Panagiotis Bamidis, and Mamas Theodorou. The legal aspect of interoperability of cross border electronic health services: A study of the european and national legal framework. *Health Informatics Journal*, 28(3):14604582221128722, 2022.
- [17] Dimitrios G Katehakis and Angelina Kouroubali. A framework for ehealth interoperability management. *Journal of Strategic Innovation and Sustainability*, 14(5):51–61, 2019.
- [18] Comisión Europea. Decisión no 1719/1999/ce del parlamento europeo y del consejo de 12 de julio de 1999 sobre un conjunto de orientaciones, entre las que figura la identificación de los proyectos de interés común, relativo a redes transeuropeas destinadas al intercambio electrónico de datos entre administraciones (ida). Technical report, Comisión Europea, 1999. URL <https://www.boe.es/doue/1999/203/L00001-00008.pdf>.
- [19] Kécia Souza Santana Santos, Larissa Barbosa Leoncio Pinheiro, and Rita Suzana Pitangueira Maciel. Interoperability types classifications: A tertiary study. 2021. doi: 10.1145/3466933.3466952. URL <https://doi.org/10.1145/3466933.3466952>.
- [20] Gabriel da Silva Serapião Leal, Wided Guédria, and Hervé Panetto. Interoperability assessment: A systematic literature review. *Computers in Industry*, 106:111–132, 2019.
- [21] Rebeca C Motta, Káthia M de Oliveira, and Guilherme H Travassos. A conceptual perspective on interoperability in context-aware software systems. *Information and Software Technology*, 114:231–257, 2019.
- [22] ACTIVIDADES SEIS. Xxi foro de seguridad y protección de datos 2024-14/02/24- tercera sesión debate, feb 2024. URL <https://www.youtube.com/watch?v=x79UKXCh1V8>.
- [23] ACTIVIDADES SEIS. Xxi foro de seguridad y protección de datos 2024-15/02/24- octava sesión, feb 2024. URL <https://www.youtube.com/watch?v=6vbbgR7MUqA>.
- [24] OHDSI. Darwin eu initiative presentation, 2023. URL <https://www.ohdsi.org/darwin-eu-initiative-presentation/>.
- [25] DigitalHealthEurope. eHDSI - European Health Data Space, 2023. URL <https://digitalhealtheurope.eu/glossary/ehdsi/>.
- [26] European Genomic Data Infrastructure (GDI) project. European genomic data infrastructure (gdi) project, 2022. URL <https://gdi.onemilliongenomes.eu/>.
- [27] vallealonsodc. Thesis-ATLAS-OHDSI. <https://github.com/vallalonsodc/Thesis-ATLAS-OHDSI>, 2024.

- [28] Junta de Andalucía. Temas: Perfiles de contratante. [https://www.juntadeandalucia.es/haciendayadministracionpublica/apl/pdc\\_sirec/perfiles-licitaciones/consultas-preliminares/detalle.jsf?idExpediente=000000078484](https://www.juntadeandalucia.es/haciendayadministracionpublica/apl/pdc_sirec/perfiles-licitaciones/consultas-preliminares/detalle.jsf?idExpediente=000000078484), 2018.
- [29] Microsoft. Comprar windows 11 pro — microsoft store españa. <https://www.microsoft.com/es-es/d/windows-11-pro>, .
- [30] Sparx Systems. Enterprise architect pricing. URL <https://sparxsystems.com/products/ea/shop/>.
- [31] Microsoft. Compra microsoft 365 personal — microsoft store españa. <https://www.microsoft.com/es-es/microsoft-365/p/microsoft-365-personal>, .
- [32] MJ Escalona, L García, JA García-García, G López-Nicolás, and N Koch. Choose your preferred life cycle and sofia will do the rest. In *Proceedings of the 23rd International Conference on Web Engineering*, 2023.
- [33] Observational Health Data Sciences and Informatics. Publications - ohdsi.org. <https://www.ohdsi.org/publications/>, .
- [34] Observational Health Data Sciences and Informatics. OHDSI youtube channel. <https://www.youtube.com/@OHDSI/playlists>, .
- [35] OHDSI discord server invitation. <https://discord.com/invite/xABFWShJYx>, .
- [36] Formulario de Microsoft Office. [https://forms.office.com/Pages/ResponsePage.aspx?id=1AAPoyCRq0q6TOVQkC0y1ZyG6Ud\\_r2tKuS0HcGnqiQZUQ05MOU9BSzEw0ThZVjNQVVFGTDNZRENONiQlQCN0PWcu](https://forms.office.com/Pages/ResponsePage.aspx?id=1AAPoyCRq0q6TOVQkC0y1ZyG6Ud_r2tKuS0HcGnqiQZUQ05MOU9BSzEw0ThZVjNQVVFGTDNZRENONiQlQCN0PWcu), .
- [37] Observational Health Data Sciences and Informatics (OHDSI). Ohdsi github repository, 2023. URL <https://github.com/OHDSI/>.
- [38] OHDSI. Ohdsi forums, 2024. URL <https://forums.ohdsi.org/>.
- [39] P. E. et al Stang. Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership, 2010.
- [40] J. M. et al Overhage. Validation of a common data model for active safety surveillance research., 2012.
- [41] A. et al Jagesar. The dutch icu data warehouse: towards a standardized multicenter electronic health record database, 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/07/3-ICUdata\\_poster\\_portrait-3-A-Jagesar.png](https://www.ohdsi.org/wp-content/uploads/2023/07/3-ICUdata_poster_portrait-3-A-Jagesar.png).
- [42] A. et al Finster. Integrating the omop cdm into the ai sandbox of the german health data lab, 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/07/7-Finster\\_OMOP-at-the-HDL\\_Poster\\_2023Symposium-Me-Li.png](https://www.ohdsi.org/wp-content/uploads/2023/07/7-Finster_OMOP-at-the-HDL_Poster_2023Symposium-Me-Li.png).
- [43] G. et al Collumeau. Standardization of the french national database snds in omop-cdm, 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/07/12-Standardization\\_of\\_SNDS\\_Health\\_Data\\_Hub-Gaelle-Collumeau.png](https://www.ohdsi.org/wp-content/uploads/2023/07/12-Standardization_of_SNDS_Health_Data_Hub-Gaelle-Collumeau.png).

- [44] L. et al Halvorsen. The laisdar project – hospital ehr harmonization in rwanda through mapping to omop cdm; outcome, challenges and lessons learned, 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/07/13-halvorsen\\_laisdarstatusposter\\_2023symposium-Lars-Halvorsen.png](https://www.ohdsi.org/wp-content/uploads/2023/07/13-halvorsen_laisdarstatusposter_2023symposium-Lars-Halvorsen.png).
- [45] M. et al van Swieten. Standardizing european sarcoma registry data to the omop common data model, 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/07/15-vanSwieten\\_Blueberry-OMOP-mapping-2023symposium-Maaike-van-Swieten.png](https://www.ohdsi.org/wp-content/uploads/2023/07/15-vanSwieten_Blueberry-OMOP-mapping-2023symposium-Maaike-van-Swieten.png).
- [46] OHDSI. Athena, . URL <https://athena.ohdsi.org/search-terms/terms>.
- [47] EHDEN Consortium. Ehden consortium, 2024. URL <https://www.ehden.eu/>.
- [48] OHDSI. Atlas demo, . URL <https://atlas-demo.ohdsi.org/>.
- [49] OHDSI github. Broadsea, 2023. URL <https://github.com/OHDSI/Broadsea>.
- [50] OHDSI github. Ohdsi in a box, . URL <https://github.com/OHDSI/OHDSI-in-a-Box>.
- [51] OHDSI github. Ohdsi aws, . URL <https://github.com/OHDSI/OHDSIonAWS>.
- [52] OHDSI github. Ohdsi on azure, . URL <https://github.com/microsoft/OHDSIonAzure>.
- [53] OHDSI github. Achilles, . URL <https://github.com/OHDSI/Achilles>.
- [54] OHDSI github. Athena, . URL <https://github.com/OHDSI/Athena>.
- [55] OHDSI github. Hades, . URL <https://github.com/OHDSI/Hades>.
- [56] OHDSI. Software tools, . URL <https://www.ohdsi.org/software-tools/>.
- [57] OHDSI github. Data quality dashboard, . URL <https://github.com/OHDSI/DataQualityDashboard>.
- [58] Ajit et al Londhi. Broadsea 3.0: “broadening the ohdsea”. 2023. URL [https://www.ohdsi.org/wp-content/uploads/2023/10/Londhe-Ajit\\_Broadsea-3.0-BROADening-the-ohdSEA\\_2023symposium-Ajit-Londhe.pdf](https://www.ohdsi.org/wp-content/uploads/2023/10/Londhe-Ajit_Broadsea-3.0-BROADening-the-ohdSEA_2023symposium-Ajit-Londhe.pdf).
- [59] Google chrome, 2024. URL [https://www.google.com/intl/es\\_es/chrome/](https://www.google.com/intl/es_es/chrome/).
- [60] Docker. Descripción general de docker, 2024. URL <https://docs.docker.com/get-started/overview/>.
- [61] PostgreSQL. Acerca de postgresql, 2024. URL <https://www.postgresql.org/about/>.
- [62] Wikipedia. Github, 2024. URL <https://en.wikipedia.org/wiki/GitHub>.

# A. Manual de ATLAS Broadsea

---

El nombre completo de este anexo corresponde a **Manual de instalación, despliegue y configuración de ATLAS Broadsea**, aunque por motivos de extensión se ha reducido en el índice de la memoria a *Manual de ATLAS Broadsea*.

El manual se presenta a la convocatoria como un documento aparte debido a su larga extensión, de casi 40 páginas. No obstante, se utiliza este apartado de la memoria para presentar resumidamente sus contenidos básicos y cómo acceder a él. Su gran extensión se debe a que recopila en un único lugar una grandísima variedad de información que hasta ahora se encontraba esparcida de forma más o menos ordenada en la red, sobretodo en diferentes repositorios de github.

El anexo se adjunta a la documentación entregable de la convocatoria con el nombre "Anexo A - Manual de ATLAS Broadsea.pdf". Adicionalmente, también es accesible a través del repositorio de github del Trabajo Fin de Grado [27], concretamente en la ruta Thesis-ATLAS-OHDSI/documentation/pdf.

El manual trata cinco aspectos importantes de ATLAS Broadsea:

1. **Introducción y descripción de Broadsea.** Este capítulo explica contenidos sobre el entorno tecnológico necesario para seguir correctamente los procedimientos del manual.
2. **Despliegue por defecto.** Este capítulo presenta el despliegue más sencillo del entorno Broadsea, sin ningún tipo de configuración adicional.
3. **Conexión con la BD por defecto.** Este capítulo explica la conexión con el servidor Postgre del contenedor docker de Broadsea.
4. **Conexión con BD externa.** Este capítulo explica cómo añadir una conexión de una base de datos externa al servidor docker de Broadsea.
5. **Configuración del Vocabulario.** Este capítulo explica cómo configurar el Vocabulario desde ATHENA y se presentan otras configuraciones avanzadas.

Todo ello complementa la información del TFG de forma subyacente, es decir, durante la reproducción del estudio práctico (véase 8) se da por supuesto todo el proceso de instalación de la herramienta así como la configuración del servidor, base de datos, etc. En términos de roles del proyecto (véase 3) se podría decir que mientras que el analista se encarga de reproducir el estudio haciendo uso de la interfaz de usuario de ATLAS, el developer habría sido el encargado de realizar toda el anexo, con toda la instalación, despliegue y configuración para que la herramienta funcione. No obstante, en este caso ambos roles son ejecutados por la misma persona que es la alumna. Además satisface explícitamente el **Obj-002** del Trabajo Fin de Grado (véase 2).

## B. Glosario

---

**Aprendizaje automático (*Machine Learning, ML*):** Campo de la inteligencia artificial que desarrolla algoritmos y modelos que permiten a las máquinas aprender a partir de datos, identificar patrones y tomar decisiones sin necesidad de ser programadas explícitamente para cada tarea específica.

**ATLAS:** Herramienta de código abierto desarrollada por la colaboración Observational Health Data Sciences and Informatics (OHDSI), diseñada para la visualización, exploración y análisis de datos de salud provenientes de diferentes fuentes y estándares, facilitando la investigación en salud pública y la toma de decisiones clínicas basadas en evidencia.

**Contenedor Docker (*Docker container*):** Tecnología de virtualización que permite empaquetar y ejecutar aplicaciones y sus dependencias en entornos aislados, proporcionando portabilidad, rapidez y consistencia en el despliegue de aplicaciones en diferentes sistemas operativos y entornos de ejecución.

**Código abierto (*Open source*):** Modelo de desarrollo de software que promueve el acceso abierto al código fuente de un programa, permitiendo su estudio, modificación y distribución por parte de la comunidad de desarrolladores, lo que fomenta la colaboración, la transparencia y la innovación en el desarrollo de software.

**Computación en la Nube (*Cloud Computing*):** Modelo de prestación de servicios de computación a través de internet, donde los recursos como almacenamiento, servidores y aplicaciones son proporcionados y gestionados por proveedores externos, permitiendo un acceso flexible y escalable según la demanda del usuario.

**Cohorte (*Cohort*):** Grupo de individuos que comparten una característica común o que han sido seleccionados para participar en un estudio de investigación, con el fin de observar y analizar los resultados de un evento o exposición específica durante un período de tiempo determinado.

**Datos masivos (*Big Data*):** Conjunto de datos extremadamente grandes y complejos que requieren tecnologías especializadas para su almacenamiento, procesamiento y análisis, con el objetivo de extraer información significativa y tomar decisiones informadas.

**Datos del mundo real (*Real World Data, RWD*):** Información sobre la salud y los resultados de atención médica recopilada de fuentes del mundo real, como registros médicos electrónicos, reclamaciones de seguros y dispositivos portátiles, utilizada para complementar los datos de ensayos clínicos y proporcionar información sobre la efectividad y seguridad de tratamientos en condiciones reales fuera del entorno controlado de un estudio clínico.

**European Health Data & Evidence Network (EHDEN):** Consorcio europeo que tiene como objetivo establecer una infraestructura escalable y sostenible para el

análisis de datos de salud del mundo real en Europa. EHDEN promueve la estandarización de datos y el uso de herramientas y métodos avanzados para facilitar la investigación clínica y epidemiológica.

**Historial Clínico Electrónico (HCE):** Registro digitalizado y centralizado de toda la información médica de un paciente, que incluye datos como diagnósticos, tratamientos, resultados de pruebas, alergias y antecedentes médicos, accesible por profesionales de la salud autorizados para mejorar la coordinación de la atención, la precisión diagnóstica y la seguridad del paciente.

**Industria 4.0 (*Industry 4.0*):** Concepto acuñado por el gobierno alemán en 2011 para referirse a la emergente cuarta revolución industrial basada fundamentalmente en la integración de los sistemas físicos con Internet a través de herramientas como Internet de las cosas, Big Data, Cloud Computing o Inteligencia Artificial.

**Internet de las cosas (*Internet of Things, IoT*):** Red de dispositivos, sistemas y servicios que incorporan sensores, software y otras tecnologías que permiten la conectividad avanzada y el intercambio de datos entre sí a través de Internet u otras redes de comunicación.

**Inteligencia Artificial (*Artificial Intelligence, AI*):** Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.

**Interoperabilidad:** Capacidad de sistemas, dispositivos o aplicaciones para intercambiar datos y trabajar juntos de manera efectiva, garantizando que la información sea comprensible y utilizada de manera consistente entre diferentes plataformas, organizaciones o entornos. Se puede clasificar en tres grupos: semántica, técnica y organizacional.

**Low-code:** Enfoque de desarrollo de software que utiliza herramientas visuales y abstracciones de código para permitir a los usuarios crear aplicaciones de manera rápida y con menos necesidad de programación manual, acelerando el proceso de desarrollo y permitiendo a usuarios con menos experiencia técnica participar en la creación de aplicaciones.

**Modelo de Datos Común de OMOP (*OMOP Common Data Model, OMOP CDM*):** Estructura estandarizada de base de datos desarrollada por la colaboración Observational Medical Outcomes Partnership (OMOP), diseñada para representar datos de salud de manera uniforme y compatible, facilitando el análisis comparativo de datos clínicos y epidemiológicos provenientes de diferentes fuentes y sistemas de salud.

**Observational Medical Outcomes Partnership (OMOP):** Iniciativa colaborativa entre la industria, académicos y reguladores para mejorar la evaluación de medicamentos a través del análisis de datos de salud del mundo real. OMOP desarrolla métodos y estándares para el análisis de datos de salud, incluido el Modelo de Datos Común (CDM), que permite la armonización de datos para la investigación.

**Omopizar:** Proceso de transformar datos de salud de diferentes fuentes y formatos al Modelo de Datos Común de OMOP (CDM), para estandarizar la representación

de los datos y facilitar su análisis comparativo y la generación de evidencia científica en investigación clínica.

**Observational Health Data Sciences and Informatics (OHDSI):** Organización internacional que desarrolla y aplica métodos de análisis de datos de salud para generar evidencia a partir de datos del mundo real, con el objetivo de mejorar la toma de decisiones en salud pública y clínica, promoviendo el uso de estándares y herramientas abiertas para el intercambio y análisis de datos.

**Salud digital (e-Salud):** Utilización de tecnologías de la información y comunicación en el ámbito de la salud para mejorar la eficiencia, accesibilidad, calidad y seguridad de los servicios médicos, así como para fomentar la participación activa de los pacientes en su cuidado y la gestión de su salud.

**Sistemas ciber-físicos (Cyber-Physical Systems, CPS):** Sistemas que integran componentes físicos y computacionales, conectados a través de redes, para monitorear y controlar procesos físicos en tiempo real, utilizando tecnologías como sensores, actuadores, y sistemas de información y comunicación.

**Sanidad 4.0 (Healthcare 4.0):** También conocido como Salud 4.0, es la aplicación de tecnologías digitales como inteligencia artificial, Internet de las cosas y big data en el sector de la salud para mejorar la atención médica, la gestión de datos y la experiencia del paciente.

**Tecnologías de la Información y Comunicación (TICs):** Conjunto de herramientas, recursos y sistemas tecnológicos utilizados para adquirir, almacenar, procesar, transmitir y presentar información de manera digital, facilitando la comunicación y el intercambio de datos entre personas, organizaciones y dispositivos.

**Telemedicina:** Práctica médica que utiliza tecnologías de la información y comunicación para realizar consultas médicas, diagnósticos, tratamiento y seguimiento de pacientes a distancia, facilitando el acceso a la atención médica y la colaboración entre profesionales de la salud sin necesidad de encuentros físicos.