

**Proyecto EHDEN: COMPARTICIÓN EUROPEA DE DATOS DE SALUD  
para desarrollar nuevas terapias y dispositivos médicos**

## Contenido

Proyecto EHDEN: COMPARTICIÓN EUROPEA DE DATOS DE SALUD para desarrollar nuevas terapias y dispositivos médicos.....	1
1. Descripción del proyecto.....	3
1.1.1. Análisis de la fuente de datos.....	4
1.1.2. Documentación para implementación de ETL .....	4
1.1.3. Implementación de la ETL .....	4
1.1.4. Pruebas.....	4
1.1.5. Documentación final y certificación.....	4
2. Repercusión para el ciudadano y las Administraciones .....	5
3. Equipo de desarrollo y proveedores .....	6
4. Valoración económica .....	6
5. Plazos de cumplimiento.....	6
6. Facilidad de uso .....	6

## 1. Descripción del proyecto

El impulso y la ingente cantidad de datos que los sistemas de información sanitarios generan y albergan son el pozo de petróleo de los datos de la vida real sanitaria de millones de personas en toda Europa. El problema surge cuanto no es posible el acceso a los mismos por falta de estandarización y vocabulario común compartido entre los diferentes servicios de salud y hospitales. Para poder tener acceso a ellos de una manera eficiente y homogénea se hace necesaria su extracción con mecanismos que garanticen la protección de datos pero que permitan abrirlos a la comunidad científica para la generación de resultados.

A finales de 2018, la Comisión Europea impulsa la creación de la Red Europea de Datos y Evidencia de la Salud (EHDEN - <https://www.ehden.eu/>) para fomentar el uso del modelo común de datos OMOP CDM en el ámbito sanitario. Los objetivos de EHDEN son ayudar a los pacientes, médicos, pagadores, reguladores, gobiernos y la industria farmacéutica a comprender la enfermedad, los tratamientos y resultados, así como a desarrollar nuevas terapias y dispositivos médicos. A la red EHDEN se han unido más de 170 organizaciones sanitarias europeas de 26 países diferentes.

Las aspiraciones de EHDEN son significativas, con el objetivo de armonizar 400 millones de registros de salud anonimizados durante cinco años correspondientes a 80 millones de pacientes en múltiples hospitales, redes de atención primaria, bases de datos regionales y similares, mientras se desarrolla una red duradera y sostenible de intereses de investigación mutuos. Esta red federada de datos (los datos no se centralizan una base de datos común, sino que cada socio los custodia en sus instalaciones) incorpora políticas y requisitos reglamentarios apropiados, como el Reglamento General de Protección de Datos (GDPR) y pautas de investigación ética, a través de la privacidad por diseño en su infraestructura.

En el año 2020, IDIVAL y el Servicio Cántabro de Salud presentan su candidatura para optar a formar parte de la red EHDEN que financia con hasta 100.000€, los trabajos e infraestructura necesaria para transformar sus datos sanitarios al modelo OMOP CDM. Para la certificación de los trabajos EHDEN exigía contar con una empresa certificada por ello en el modelo OMOP CDM. IDIVAL y el SCS contaron para ello con Bahía Software. Bahía Software, con amplia experiencia en sus más de 23 años de trayectoria en la informática sanitaria, es una de las empresas certificadas por EHDEN para la armonización de los datos OMOP y ha desarrollado una serie de herramientas que aceleran y optimizan la conversión de los datos. Ello favorece en gran medida la compartición de los mismos entre las diferentes organizaciones sanitarias. Esta colaboración es la que Bahía Software viene prestando a instituciones como el la Fundación Rioja Salud, el Instituto Aragonés de Ciencias de la Salud, el Parc Sanitari Sant Joan de Déu, el Consorci Corporació Sanitària Parc Taulí, el Departament de Salut de Catalunya, el Institut Català d'Oncologia y el Centro Clínico Académico de Braga (Portugal) donde se han gestionado los datos de más de 6 millones de pacientes y que llegará a una cifra superior a los 27 millones con la incorporación de nuevas organizaciones.

La planificación de las fases del proyecto fue la siguiente:

### **1.1.1. Análisis de la fuente de datos**

Se trata de una tarea que requirió la involucración de personal con conocimientos detallados de los datasets origen, personal de IDIVAL/SCS y de los proveedores de los sistemas de información (Altamira, LIS, RIS, Atención Primaria, etc). El equipo de proyecto obtuvo una visión global de la información almacenada y la que se quería transformar.

Se celebraron sesiones para entrar poco a poco en el detalle de las tablas implicadas, sistemas de codificación, volumen de datos y demás características a tener en cuenta de cara al mapeo a la estructura OMOP.

### **1.1.2. Documentación para implementación de ETL**

Es la segunda tarea que requirió también la involucración de personal con conocimientos detallados de los datasets origen. Se celebraron varias sesiones de trabajo conjunto para seguir las instrucciones necesarias para el uso de las herramientas WhiteRabbit y Rabbit-in-a-hat.

- WhiteRabbit: Permitía extraer reporte de las tablas, limpiar valores de columnas que no se quisieran compartir y traducir los nombres de columnas.
- Rabbit-in-a-hat: Permitía cruzar tablas y campos del modelo origen al modelo CDM. Añadir todos los detalles posibles y las transformaciones necesarias sobre cada uno de los campos a mapear.

El resultado final fue el documento final de mapeo con información necesaria para la implementación de la ETL por parte de cualquier técnico sin conocimientos funcionales sobre los datasets origen.

### **1.1.3. Implementación de la ETL**

Esta tarea tuvo como punto de partida el documento generado en el paso anterior con Rabbit-in-a-hat. Se hizo uso de Usagi para el mapeo de conceptos que pertenecían a sistemas de codificación que no estuvieran contemplados en OMOP CDM.

Finalmente, se empleó una herramienta específica para la construcción de las ETL y volcado de datos a la BD destinada a almacenar la información basado en el modelo de datos OMOP.

### **1.1.4. Pruebas**

Se realizaron los tests necesarios sobre la ETL construida en el paso anterior de cara a comprobar su correcto funcionamiento además de evaluar la calidad de los datos.

### **1.1.5. Documentación final y certificación**

Bahía Software acompañó al personal de IDIVAL y/o del Servicio Cántabro de Salud en las sesiones de seguimiento con el equipo de trabajo europeo EHDEN. Ello incluso la elaboración del informe final y envío de las evidencias que permitió la certificación de la base de datos OMOP por parte de EHDEN.

La base de datos OMOP tiene un tamaño final de más de 90 Gb de información como resultado de la armonización de la siguiente cantidad de datos procedentes de los diversos sistemas:

- Altamira: 27Gb
- CMBD: 139Mb
- GAP: 59Gb
- PEA: 793Mb

- RE: 905Mb
- TICARES: 4,3Gb

## 2. Repercusión para el ciudadano y las Administraciones

Como hemos venido aprendiendo en los últimos años a raíz de la aprobación de diferentes vacunas para la COVID19, los ensayos clínicos son el método por excelencia para generar evidencia científica. Sin embargo, cada vez más se está impulsando el empleo de los datos ya generados por las instituciones sanitarias para realizar descubrimientos empleando técnicas de Inteligencia Artificial, Big Data y Machine Learning. Se trata de los conceptos Real World Data (RWD) y Real World Evidence (RWE). En otras palabras, se trata de la recogida de datos sobre medicamentos, pacientes y enfermedades en la práctica clínica diaria.

Que IDIVAL y el SCS cuenten con los datos sanitarios armonizados en el modelo OMOP CMD supone una serie de ventajas tanto para los ciudadanos como para las administraciones.

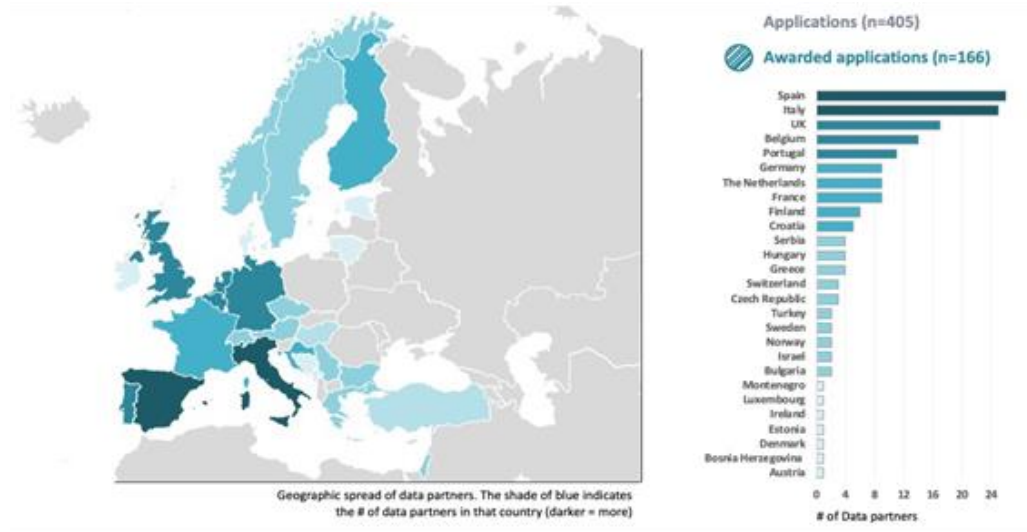
Para los ciudadanos, las ventajas son las siguientes:

- Mejora de la calidad y la accesibilidad de los datos sanitarios: El modelo OMOP CDM proporciona un marco común y estandarizado para la organización de los datos sanitarios, lo que facilita su análisis y comprensión. Esto permite a los profesionales sanitarios proporcionar una atención más personalizada y efectiva a los pacientes.
- Mayor transparencia y participación: El modelo OMOP CDM facilita la publicación de los datos sanitarios de forma abierta y accesible, lo que permite a los ciudadanos conocer mejor su estado de salud y participar en las decisiones sobre su atención.
- Mejora de la investigación sanitaria: El modelo OMOP CDM facilita la integración de datos de diferentes fuentes, lo que permite a los investigadores realizar estudios más completos y precisos. Esto contribuye a la mejora de la prevención, el diagnóstico y el tratamiento de las enfermedades.

Para las administraciones, las ventajas son las siguientes:

- Mayor eficiencia y eficacia en la gestión de los recursos sanitarios: El modelo OMOP CDM permite a las administraciones sanitarias optimizar la asignación de recursos, lo que contribuye a mejorar la calidad y la eficiencia de la atención sanitaria.
- Mejora de la planificación y la evaluación de las políticas sanitarias: El modelo OMOP CDM facilita la recopilación y el análisis de datos sanitarios para la planificación y la evaluación de las políticas sanitarias. Esto permite a las administraciones tomar decisiones más informadas y efectivas.

Mayor colaboración entre las administraciones sanitarias: El modelo OMOP CDM facilita la colaboración entre las diferentes administraciones sanitarias, lo que contribuye a mejorar la equidad y la eficiencia del sistema sanitario.



### 3. Equipo de desarrollo y proveedores

El equipo de proyecto por parte del IDIVAL/SCS estuvo formado por profesionales con un amplio conocimiento de los sistemas de información de la organización. Su función fue crucial para garantizar una adecuada comunicación entre los diferentes actores (equipos técnicos de los proveedores y de Bahía Software) así como de garantizar en todo momento el respeto por la privacidad y seguridad de los datos de los pacientes.

En el proyecto participaron además equipos de trabajo mencionados, personal de los siguientes sistemas:

- Altamira – Historia Clínica Electrónica
- CMBD - Conjunto Mínimo Básico de Datos
- GAP – Gerencia de Atención Primaria
- PEA – Prescripción Electrónica Asistida
- RE – Receta Electrónica
- TICARES – Sistema de Citación

### 4. Valoración económica

La financiación aportada por EHDEN a IDIVAL/SCS ascendió a 100.000€

### 5. Plazos de cumplimiento

El proyecto se inició el octubre de 2020 y finalizó con la certificación de la base de datos OMOP en diciembre de 2021.

### 6. Facilidad de uso

El uso de OMOP supone ventajas y facilidades en investigación clínica por los siguientes motivos:

- OMOP es un modelo de datos estandarizado y abierto, lo que facilita la comprensión y el uso de los datos por parte de los investigadores. El modelo OMOP define un conjunto de tablas y campos comunes para la representación de los datos sanitarios, lo que permite a los investigadores centrarse en el análisis de los datos sin tener que preocuparse por la estructura de los mismos.
- OMOP y su ecosistema OHDSI, proporciona una serie de herramientas y recursos para facilitar la investigación clínica en formato de software open source que aportan una amplia gama de recursos para ayudar a los investigadores a trabajar con datos OMOP, incluyendo guías, tutoriales y ejemplos de código.
- OMOP es compatible con una amplia gama de herramientas y software de investigación. La mayoría de las herramientas y software de investigación sanitaria actuales son compatibles con datos OMOP, lo que facilita a los investigadores el acceso a los datos y el análisis de los mismos.

En concreto, la facilidad de uso de OMOP para la investigación clínica se puede ver reflejada en los siguientes aspectos:

- Reducción del tiempo de preparación de los datos. El modelo OMOP estandariza la estructura de los datos, lo que reduce el tiempo necesario para preparar los datos para su análisis.
- Mejora de la calidad de los datos. El modelo OMOP proporciona un conjunto de reglas y recomendaciones para la codificación de los datos, lo que ayuda a mejorar la calidad de los mismos.
- Mayor colaboración entre investigadores. El modelo OMOP facilita la colaboración entre investigadores, ya que permite compartir datos de forma estandarizada.