

UIMP 2024. ENCUENTRO ISCIII

CCSHARED. Un caso de uso

Our partners



**IDI
BELL**

Bellvitge Biomedical
Research Institute



**IDI
BELL**

Bellvitge Biomedical
Research Institute

The campus



CCShared-CC

Federated population health data: intelligent identification of shared factors in Cancer and Cardiovascular diseases to improve prevention and disease management across the lifespan

1

SUBPROJECT 1 & COORDINATOR: VICTOR MORENO. Coordinacion y federacion de datos para un analitica distribuida (dataFED)

2

SUBPROJECT 2: JOSEP COMIN. Estrategias preventivas integradoras impulsadas por inteligencia artificial (IA) para el manejo de cancer y las enfermedades cardiovasculares (ARTICCA)

3

SUBPROJECT 3: JOSEP MUNUERA. Herramientas de evaluacion del riesgo cardiovascular en pacientes pediátricos oncologicos basadas en inteligencia artificial (IA) (AI_CARONTe)

4

SUBPROJECT 4: PIs: GABRIEL CAPELLA & ERNEST NADAL. Desarrollo de una herramienta de ayuda para la toma de decisiones y gestión multidisciplinar del cuidado del cancer basada en Inteligencia Artificial (AID-CANCER CARE)

Objectives

Overarching goal

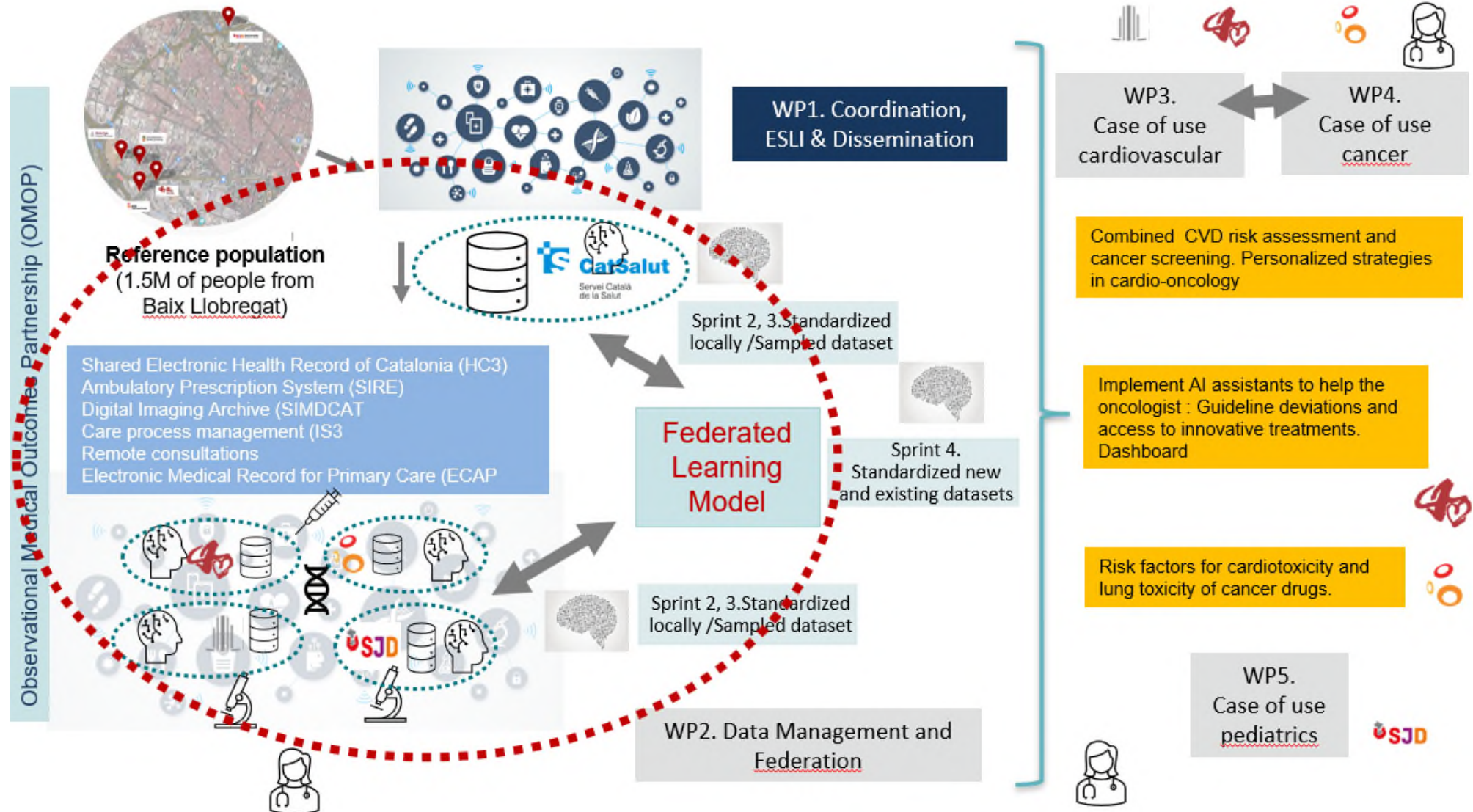
To develop models to improve the stratification of the population based on shared risk factors for the prevention and treatment of pleiotropic diseases such as cancer and cardiovascular diseases, throughout life, using an integrated analysis of multidimensional data gathered from multiple sources and exploited in a federated global learning model.

Objectives

Main goals

1. To **digitally transform all data** related to the working hypothesis (clinical, genetic, histopathological, radiological, and socio-environmental) and **to create a federated global learning model**, that uses data integration and harmonization at each site as the underlying framework (PI V Moreno)
2. To **develop and validate shared cardiovascular-cancer risk prediction models** using AI/ML approaches in the Federated Global Learning Model.
3. To **identify risk factors for cardiotoxicity of cancer drugs**, both in the onco-hematologic pediatric population and in the adult population.
4. To **develop AI-based assistants** to help the clinicians involved in breast and lung cancer care to reach decisions in the context of multidisciplinary tumor boards. (PI: G Capellá, E Nadal)

Objectivos



To **digitally transform all data related to the working hypothesis** (clinical, genetic, histopathological, radiological, and socio-environmental)

To **create a federated global learning model**, that uses data integration and harmonization at each site as the underlying framework (PI V Moreno)

Aprobación por parte del CEIm:

Protocolo master aprobado. Protocolos específicos para cada estudio. Evaluado con gran detalle: algoritmos de inteligencia artificial, compartición de los datos

Inventario de datos en 3 instituciones: hay muchos datos fragmentados (no todo está en SAP)

Datos mapeados en OMOP (cuesta (mucho) dinero)

ICO: todos los pacientes

HUB: cardiovascular

SJD: todos los pacientes

Trabajando con una solución DataSHIELD (en R) para el análisis de datos federados (open source)

To develop and validate shared cardiovascular-cancer risk prediction models using AI/ML approaches in the Federated Global Learning Model

- Protocolo aprobado
- Definición de la cohorte completada
- Datos definidos por pacientes: análisis vertical (split entre HUB (Cardio) e ICO (Onco); mismo paciente en dos sistemas)
- 35000 pacientes en OMOP
- Es necesario compartir datos hasta que el aprendizaje federado esté en marcha

Comentarios

- Tiempo prolongado para la obtención de las aprobaciones éticas
- El mapeo al formato OMOP es un proceso que necesita mucho tiempo (diferentes fuentes con diferentes formatos)
- El proceso de federación de datos necesita un recurso informático potente
- La comunicación IT entre entidades es importanted
- La encriptación de los datos por parte de CatSalut
- Convenio de gobernanza de datos que:
 - defina las modalidades de acceso a los datos entre las diferentes entidades en futuros proyecto de compartición de datos.
 - Mitigue las muchas reticencias a compartir datos (hospitales son los responsables del tratamiento de datos del paciente) dejando poco margen a la interpretación
 - Sea compatible con la normativa europea que quieren aplicar. El Reglamento de PD es muy garantista con el sujeto fuente.

Comentario final: ¿qué es un dato anónimo?

anónimo *per se*

VS

en ninguna circunstancia podrá ser identificable

<https://datos.gob.es/ca/blog/especificaciones-une-gobierno-gestion-y-calidad-del-dato>

datos.gob.es

REGULACIÓN RELATIVA A LOS DATOS ABIERTOS Y LA REUTILIZACIÓN DE LA INFORMACIÓN DEL SECTOR PÚBLICO EN ESPAÑA

Contenidos clave de la Ley 37/2007 sobre reutilización de la información del sector público, tras la incorporación de las previsiones de la Directiva (UE) 2019/1024 relativa a los datos abiertos y la reutilización de la información del sector público.



OBJETIVO:



Autorización generalizada de reutilización de los datos del sector público aplicando el principio de datos abiertos desde el diseño y por defecto.

ÁMBITO DE APLICACIÓN



- Datos producidos por la Administración General del Estado, las Comunidades Autónomas y entidades que integran la Administración Local.
- Datos generados por organismos y entidades del sector público institucional, incluidas bibliotecas, museos y archivos.
- Datos y servicios referidos en la Directiva INSPIRE.
- Datos de investigación financiada con fondos públicos.
- Datos de determinadas sociedades mercantiles.