ANEXO III. DESCRIPCIÓN DEL RETO

Las siguientes especificaciones de la necesidad no cubierta son orientativas. Estas especificaciones podrán evolucionar a medida que se vaya actualizando el estado del arte en cada uno de los ámbitos.

ARQUITECTURA DE DATOS DE SALUD PARA EL APRENDIZAJE (INFOBANCO)

ANTECEDENTES

El uso secundario de los datos de salud comprende la utilización de la información personal de salud para un uso distinto de la prestación de asistencia directa, e incluye actividades tales como investigación, innovación, evaluación de resultados, calidad, seguridad, epidemiología, salud pública, financiación y acreditación. Ahora y en el futuro, para un sistema sanitario en continua mejoría, "que aprende", es necesaria una atención sanitaria de calidad, capaz de expandir el conocimiento sobre la enfermedad, proveyendo los mejores tratamientos personalizados, con un aumento de la efectividad y la eficiencia, un fortalecimiento de los objetivos de seguridad de los pacientes y de la salud pública.

Para el uso secundario de la información de salud son necesarias plataformas con solvencia tecnológica, que permitan recopilar, combinar y proporcionar los datos para los diferentes objetivos, con una adecuada relevancia científica, cumpliendo los requisitos de seguridad y confidencialidad, de acuerdo con directrices éticas y que sean socialmente aceptados. La experiencia internacional de los últimos diez años ya ofrece conclusiones en este sentido, mientras proporciona estándares, herramientas y buenas prácticas que pueden ser muy útiles. Estos modelos tecnológicos y organizativos pueden servir de base para la adopción en la región de Madrid de un marco corporativo que comprenda infraestructura, personas, gobernanza y procesos de gestión de los datos de salud para el uso secundario, producidos desde varios ámbitos, que incluyen la asistencia sanitaria desde todos los niveles, incluyendo el máximo de centros del servicio regional de salud, las direcciones generales de la Consejería de Sanidad, y con la interacción de datos poblacionales, ambientales e incluso de diferentes dispositivos.

Esta estructura se puede concebir como una red regional de datos de salud (Infobanco),

con un repositorio que combina información originada en diferentes fuentes, tanto de sistemas de clínicos, de investigación y administrativos. La identificación única del sujeto de la asistencia y la cobertura poblacional en nuestro sistema de salud, proporciona datos con un nivel de exhaustividad y seguimiento muy destacados a nivel internacional. Esta red, conectaría con una red nacional y con las redes europeas, actuales y en gestación, basadas en los mismos principios. El propósito de esta red sería combinar y compartir la información con un significado preciso y bien definido, común para la interpretación homogénea de los datos, y con las medidas de seguridad y protección de privacidad requeridas.

La información de salud de estos repositorios se puede utilizar tanto en los proyectos de los centros como para los proyectos en consorcio, principalmente los europeos, proporcionado una gran ventaja a todas las candidaturas de investigadores de la región de Madrid, ya que dispondría de un repositorio con gran volumen de datos, de carácter poblacional, con seguimiento temporal e información precisa de los pacientes.

Esto es posible porque la posición del sistema regional de salud en cuanto a datos de salud es privilegiada a nivel europeo, dada la organización y el grado de digitalización de nuestro sistema de salud: cobertura poblacional basada en la atención primaria, y hospitales de referencia en su mayoría públicos, con una identificación que permite combinar los datos individuales y de forma longitudinal de la mayoría de la asistencia sanitaria. Esto representa un activo excepcional para la participación en redes de datos de salud, de ámbito nacional e internacional con diferentes grados de anonimización según la legislación aplicable.

La tecnología está disponible, ha sido probada en diferentes proyectos norteamericanos y europeos, y la plataforma regional adoptaría estas soluciones alineados con los consorcios internacionales a los que nos podríamos sumar. Como consecuencia de la adopción de estos modelos y sus herramientas, la región podría sumarse a las iniciativas internacionales que promueven su utilización y desarrollo, fundamentalmente: tranSMART y OHDSI, adquiriendo una experiencia imprescindible en estos métodos y tecnologías abiertas, participando en los consorcios internacionales permitirá el mejor aprovechamiento para las necesidades del sistema regional de salud.

En la región se desarrollan actualmente diferentes experiencias en la utilización de estos estándares y plataformas, tanto en proyectos nacionales, como europeos. En ellos se ha generado un gran avance en el conocimiento y aplicación de los estándares, tecnologías y análisis necesarios, de forma que constituyen una primera demostración piloto, que pueden

ser escalables en su implantación, incorporando otros centros y servicios.

La estructura de la plataforma en la región seguiría las experiencias realizadas con éxito en Estados Unidos: nodos en los centros, que se federan para combinar los datos de los diferentes orígenes. Los modelos de datos y las herramientas abiertas de software que se implantarían en la plataforma serían i2b2, y OMOP, con nodos en los centros e institutos de investigación conforme se vayan cubriendo los requisitos necesarios, conteniendo la información poblacional del área de referencia y con un nodo central regional que recopile toda la información de los ciudadanos. Dado que en la plataforma los datos quedarán exclusivamente dentro de la seguridad de la red regional, en los proyectos que pudieran extender el alcance de la región, se establecerían las regulaciones y procedimientos para la compartición de datos en esos niveles, como ya se contempla en estas experiencias internacionales.

Los procesos técnicos de normalización, transformación y carga de datos, los métodos de análisis y explotación, la gestión de la seguridad y confidencialidad, las políticas de acceso y compartición de datos y, en general, los diferentes elementos están desarrollados suficientemente por los promotores internacionales de estos repositorios, y se trataría de adaptarlos a nuestro entorno. La gobernanza y gestión de este "Infobanco" en red, en nuestro medio sería análoga a la de los biobancos, que están consolidados en las organizaciones sanitarias y por los investigadores, y son socialmente aceptados. Esta visión práctica de la utilización secundaria de los datos de salud refuerza la vinculación de diferentes "Infobancos", biobancos y centros de genómica, que es estratégica para el desarrollo de la investigación e innovación en salud.

ESTADO DEL ARTE

Las soluciones actuales no toman como referencia modelos de datos normalizados, ni tampoco la necesidad de interoperabilidad con sistemas externos. Esto hace que sean soluciones que solo sirven para una situación determinada, pero es imposible que se puedan generalizar a otros escenarios más ambiciosos.

La propuesta de creación de una arquitectura de datos (Infobanco) tiene como objetivo solventar esta limitación mediante la creación de redes que definen un modelo de información y vocabulario común, y un conjunto de herramientas y servicios para facilitar la

compartición de datos en la red.

Entre las herramientas y modelos comunes se encuentran:

- i2b2 y OMOP como modelos compartidos de convergencia para la persistencia y explotación de datos de salud.
- ISO/EN 13606 como modelo de referencia [3] para la construcción de arquetipos [4] para el modelado de conceptos clínicos.
- HL7 y CDISC como estándares de comunicación de datos de salud.
- Sistemas terminológicos y de clasificación para la construcción de un vocabulario de intercambio común y enlazar semánticamente los modelos.
- Tecnologías semánticas avanzadas para la transformación entre modelos de información.

El uso secundario de los datos de salud comprende la utilización de la información de salud para investigación, evaluación de resultados, calidad y seguridad, epidemiología y salud pública, financiación y acreditación. Como se enuncia en la actualidad, es necesaria para el sistema sanitario "que aprende": mejorar la atención sanitaria, expandir el conocimiento sobre la enfermedad y los tratamientos adecuados, fortalecer la comprensión sobre la efectividad y la eficiencia de los sistemas de salud, apoyar los objetivos de seguridad de pacientes y salud pública.

Para el uso secundario de la información de salud son necesarias unas plataformas con solvencia tecnológica que permitan recopilar y proporcionar los datos para los diferentes propósitos, con relevancia científica, cumpliendo los requisitos de seguridad y confidencialidad, de acuerdo a directrices éticas y que sea socialmente aceptado. La experiencia de los últimos diez años ya ofrece claramente conclusiones en este sentido y asimismo proporciona estándares, herramientas y buenas prácticas que pueden ser utilizadas.

Los repositorios de datos de salud han de cumplir una serie de requisitos como son: normalización del modelo de información clínica y terminología controlada.

En este proyecto creamos el término "Infobanco" para referirnos al repositorio de datos de salud de uso secundario, por su analogía con los biobancos como plataforma de apoyo a

la investigación, con el propósito de facilitar la difusión del concepto en nuestro medio.

Además de para la atención sanitaria y social, el uso de información sanitaria en repositorios que permitan su explotación analítica es cada vez algo más demandado. En la escena internacional, se lleva años trabajando en este campo, lo que ha dado lugar al desarrollo de políticas de actuación, modelos de repositorios y su aplicación en forma de proyectos competitivos.

En el plano nacional se ha venido trabajando de manera local, sin tomar como referencia modelos de datos normalizados y la integración con sistemas externos, teniendo una solución aplicable solo a una situación determinada, pero imposible de generalizar a otros escenarios.

En esta propuesta se pretende solventar esta limitación mediante la creación de redes que definen un modelo de datos y un vocabulario común, con un conjunto de herramientas y servicios para facilitar la compartición de datos en la red.

NECESIDADES NO CUBIERTAS

Las soluciones nacionales, no siempre toman como referencia modelos de datos normalizados, ni consideran la necesidad de interoperar con sistemas externos. Esto hace que sean soluciones que en ocasiones solo sirven para una situación determinada, pero con limitaciones para ser generalizados a otros escenarios más ambiciosos. Sin embargo, si existen soluciones en el marco europeo, que necesariamente utilizan estándares para asegurar la interoperabilidad de los datos. Los proyectos nacionales y europeos de uso secundario de datos de salud en los que contribuyen los centros de la Comunidad de Madrid representan los mayores casos de éxito hasta la fecha, y contribuyen la base del aprendizaje para el desarrollo del Infobanco en nuestro medio.

Como origen y principal experiencia son la trayectoria desde 2009 de proyectos coordinados financiados por la estrategia de salud del ISCIII sobre uso secundario de la HCE, que ha desarrollado un repositorio de datos de salud de modelo dual basado en la norma ISO 13606 con la Unidad de Telemedicina del ISCIII, demostrando la capacidad de combinar con pleno significado la información de diferentes sistemas y centros sanitarios y su uso primario y secundario.

En el proyecto europeo "Electronic Health Records for Clinical Research¨" (EHR4CR) del

instrumento Horizonte 2020 IMI se desarrolló InSite/TriNetX una plataforma para la utilización de los datos de salud basada en el aprendizaje y la mejora, consistente en un repositorio de datos sobre el modelo abierto i2b2, y una interface de interrogación y análisis de datos. A partir de esta plataforma se ha construido una red paneuropea de centros sanitarios para demostrar el valor de los Datos del Mundo Real actualizados (Real World Data RWD) para la investigación clínica y la evaluación de la atención sanitaria.

El proyecto europeo ""Electronic Health Records for Electronic Data Capture Research" (EHR2EDC) del instrumento EIT Health ha demostrado que se puede utilizar una parte importante de los registros clínicos digitales para poblar datos en los Formularios de Reporte de Casos electrónicos de la investigación clínica y epidemiológica, mediante la utilización de los estándares de datos para investigación como CDISC e i2b2.

El proyecto "European Health Data Network (EHDN)" del instrumento Horizonte 2020 IMI está construyendo una "Red Europea de Datos de Salud", basada en repositorios con el modelo OMOP y apoyada en la comunidad OHDSI, con el propósito de ser utilizados para cuestiones de investigación específicas de tipo observacional.

En los proyectos europeos de uso secundario de datos de salud se han incorporado centros sanitarios de la Comunidad de Madrid, que cumplen una serie de requisitos relacionados con el nivel de implantación de Historia Clínica electrónica y de la adopción de estándares para la normalización de los datos de salud. Además, se ha demostrado la factibilidad y utilidad de la combinación de datos de salud de poblaciones procedentes de Atención Primaria y de Hospital centrados en el paciente, con cobertura poblacional y en organizaciones que tienen continuidad asistencial. Nuestros sistemas de salud, con estructuras de aseguramiento universal y público, geográficamente organizados, con base en Atención Primaria, con hospitales por áreas y cartera de servicios de referencia, y con la identificación única del paciente, hace que sea factibles disponer de repositorios únicos en el momento actual.

El objetivo principal de los proyectos a nivel europeo es demostrar el valor de los Datos del Mundo Real actualizados (Real World Data RWD) para la investigación clínica y los estudios observacionales con la aplicación de la tecnología de repositorios de modelos normalizado a gran escala.

Las plataformas están diseñadas como primer propósito para la exploración de cohortes. A

través de una sintaxis de interrogación que utiliza los términos controlados y expresiones que restringen el tiempo y los contactos con la organización sanitaria, se identifica el número de casos que cumplen los criterios. A partir de esta se caracteriza y se pueden explorar relaciones entre variables y la dependencia con el tiempo. En el modo más avanzado abarcan datos genómicos e incluyen el análisis de asociaciones fenotipo con genotipo.

Los casos de uso comprobados de los proyectos son investigación clínica, estudios observacionales epidemiológicos y de evaluación de resultados de la asistencia sanitaria. En el proyecto EHR2EDC se ha demostrado la capacidad de poblar los sistemas de recogida de datos electrónicos para investigación y epidemiología.

El análisis y la visualización de información y la generación de conjuntos de datos, combinando los diferentes orígenes, se pueden utilizar para la atención sanitaria basada en valor y la generación de nuevo conocimiento.

OBJETIVO GENERAL

Desarrollar una arquitectura de red regional de datos de salud (Infobanco), concebido como un repositorio normalizado de datos de salud, en el que se combina información generada en diferentes fuentes, tanto sistemas de clínicos, administrativos y de investigación. Esta arquitectura tendrá un funcionamiento de plataforma que provea servicios a clínicos, gestores e investigadores, y estará dotada de herramientas de gobernanza, obtención, transformación, interrogación, visualización y análisis de datos para la obtención de conocimiento y el soporte a la toma de decisiones.

OBJETIVOS ESPECÍFICOS

- Diseñar una arquitectura funcional y tecnológica, basada en el estado del arte, que permita responder a los objetivos planteados.
- Implementar la arquitectura diseñada en uno o más hospitales del SERMAS que cumpla con los requisitos que se establezcan en el proyecto.
- Evaluar la plataforma de datos de salud para dimensiones generales de calidad: unicidad, completitud, consistencia, estabilidad multi-origen y exactitud.

- Determinar la validez de la plataforma de datos de salud humana, frente al registro manual tomado como "gold standard".
- Cuantificar la utilidad de la plataforma de datos de salud en investigación, generación de conocimiento y mejora de la asistencia.

RESULTADOS ESPERADOS

-

El INFOBANCO consistirá en una arquitectura de datos de salud y de procesos asociados de gobierno de los datos para permitir el aprendizaje basado en los datos del sistema de salud.

En tanto que arquitectura, el INFOBANCO proveerá herramientas de extracción, trasformación y carga de datos desde los distintos sistemas operacionales e informacionales implementado en los centros y servicios del sistema de salud, así como un lago de datos (Data lake) con distintas áreas de procesamiento, repositorios normalizados de uso secundario, una plataforma para la elaboración de cuadernos de recogida de datos (CRD), interfaces normalizadas para la integración con plataformas analíticas, sistemas de soporte a la decisión y de Inteligencia Artificial (IA), interfaces de interrogación y extracción de datos, herramientas de análisis y visualización de datos, herramientas de calidad de datos y herramientas para garantizar la seguridad y confidencialidad de datos, entre otras.

Adicionalmente, el INFOBANCO proveerá de un conjunto de procesos de gobierno de los datos, incluyendo normalización, intercambio, almacenado, recuperación, análisis, calidad y protección, que permitan la utilización efectiva de datos de salud y de gestión por parte de las organizaciones sanitarias.

Disponer de un INFOBANCO permitirá analizar de forma integrada datos clínicos, de gestión asistencial y económica para la mejora asistencial, la atención sanitaria basada en el valor (VBHC) y la investigación biomédica. De esta forma, la organización sanitaria dispondrá de datos, información y conocimiento para:

- La toma de decisiones de gestión, tanto en el ámbito gerencial como de los mandos intermedios y de gestión clínica de los profesionales asistenciales en su operativa diaria, potenciando el liderazgo a todos los niveles.

- La transformación de procesos asistenciales y el aprendizaje continuo, en base a los resultados obtenidos.
- El soporte a la toma de decisiones clínicas, tanto para pacientes concretos como para poblaciones de pacientes.
- El uso eficiente de los recursos y espacios.
- La gestión asistencial y la investigación sobre cohortes de pacientes y en resultados de salud.
- El conocimiento de los pacientes, la personalización y humanización de sus actuaciones incorporando la voz del paciente en la toma de decisiones, la mejora de la experiencia del paciente y la mayor implicación del paciente en la gestión de su enfermedad.
- La identificación de buenas prácticas en la organización y de las necesidades de capacitación y entrenamiento de los profesionales.
- La creación de redes de intercambio y procesamiento de datos entre organizaciones dotadas de un INFOBANCO.
- Además del desarrollo de:
    - En cuanto a Medicina Genómica: asociaciones fenotipo genotipo (pilares IMPACT) y conexión con MEDIGENOMICS.
    - Analítica e IA de conjuntos de datos de atención al envejecimiento (conexión con INTEGRACAM).
    - Algoritmos de automatización de tareas de procesos para introducir un proceso planificado, organizado y monitorizado de aprendizaje que facilite el desarrollo de la innovación local y la incorporación de nuevas capacidades.
    - Algoritmos de automatización de ayuda a la decisión clínica, que permitan avanzar en la agenda de transformación digital al servicio de la práctica diaria tanto clínica como de generación y difusión del conocimiento.
    - Cuadros de mando clínicos de proceso y condiciones de salud para identificar, priorizar y monitorizan grupos de pacientes por sus condiciones

clínicas y sus determinantes de salud e iniciar intervenciones proactivas, según las necesidades de cada grupo.

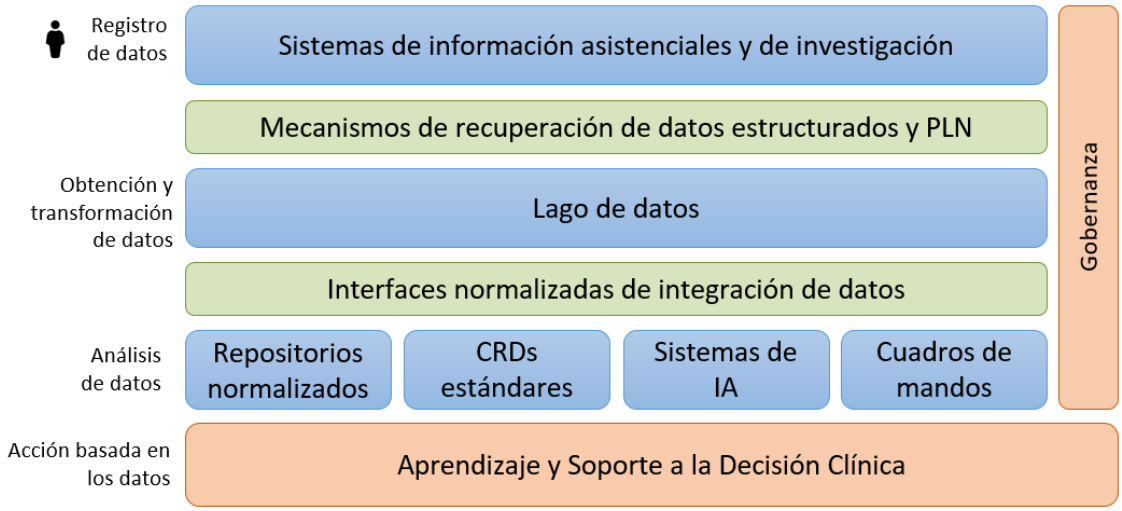
- Empoderamiento de pacientes: facilitando el trabajo de manera planificada, anticipada, y estructurada para mejorar la trayectoria asistencial de los pacientes y fomentar un mejor control de su salud y su participación activa.
- Generación de CRD para investigación clínica y epidemiología y repositorios normalizados para estudios observacionales y ensayos clínicos con RWD para general conocimiento mediante la investigación traslacional y servicios de salud para accionar la transformación asistencial.

En cuanto a la arquitectura, el INFOBANCO deberá proveer herramientas para facilitar funciones como la obtención y transformación de datos desde los sistemas asistenciales y de investigación, para el almacenamiento de los datos procesados y su análisis y para la realización de acciones sobre otros sistemas en base a los datos. Para ello, serán necesarias herramientas como:

- Herramientas de extracción, trasformación y carga (ETL) de datos desde los distintos sistemas operacionales e informacionales, entre los componentes de la arquitectura, así como desde sistemas externos.
- Herramientas para la extracción de datos estructurados a partir de textos no estructurados a través de tecnologías de procesamiento de lenguaje natural (PLN).
- Un lago de datos (Data lake) con distintas áreas de procesamiento como elemento nuclear de la arquitectura.
- Servidor de modelos (arquetipos) y de terminología (SNOMED-CT).
- Repositorios normalizados de uso secundario (RWD) basados en estándares (ISO 13606/openEHR, i2B2, OMOP) alimentados desde el lago de datos u otros que se considere en el momento.
- Una plataforma para la elaboración de cuadernos de recogida de datos (CRD), que puedan servir tanto de sistemas fuente como ser alimentados desde el lago de datos.

- Interfaces normalizadas y basadas en estándares (como por ejemplo FHIR, CDISC, etc.) para la integración con sistemas asistenciales, plataformas analíticas y de BI, sistemas de soporte a la decisión y de Inteligencia Artificial (IA), tanto para su entrenamiento como para su despliegue en producción.
- Interfaces de interrogación y extracción de datos, con conectores estándar para su consumo por terceros de una forma segura.
- Herramientas de análisis (Business Intelligence, BI) y visualización de datos.
- Herramientas de evaluación de la calidad de los datos.
- Herramientas para garantizar la seguridad y confidencialidad de datos, incluyendo cifrado, protección de autoría e integridad, anonimización y seudoanonimización.
- Entre otras que considere el mercado.

A continuación, se ilustra, a modo de ejemplo, una arquitectura tipo, no siendo esta exclusiva.



En cuanto a procesos, el INFOBANCO permitirá:

- Establecer un modelo de gobernanza sobre los modelos de información, el registro estructurado y el acceso a los datos de salud y sus múltiples usos que, entre otros aspectos, deberá contemplar:

- Gestión de estándares (terminologías, modelos, intercambio, indicadores).
- Gestión de solicitudes de utilización (proyectos, valoración ética-científica).
- Políticas de acceso de usuarios multi-organización (roles, organizaciones, grupos, permisos).
- Gestión de seguridad y riesgos (ENS, LOPDGDD, RGPD, etc.).
- Manejo de datos identificativos (anonimización y pseudoanonimización, ofuscación, etc.).
- Gestión de la propiedad intelectual (modelos, algoritmos, etc.).

- Definir un marco de convergencia a través de estándares y plataformas de referencia para la implementación de productos de información sanitaria.
- Proveer de una infraestructura segura de acceso a datos, de acuerdo a la normativa vigente y a las mejores prácticas de la industria, para su uso en ayuda a la decisión, automatización, investigación y gestión clínica.
- Definir los indicadores y procedimientos para cuantificar la calidad, validez y utilidad de los datos para estos propósitos.
- Diseñar e implementar los mecanismos de recuperación de datos estructurados y narrativos desde los sistemas fuentes.
- Diseñar e implementar las interfaces de integración entre el lago de datos y los sistemas consumidores de información, incluyendo el mapeo de modelos.
- Diseñar e implementar el conjunto básico de recursos de datos de salud para el aprendizaje (CRD, repositorios, cuadros de mando, IA).
- Facilitar una formación avanzada en el manejo de datos para el conjunto de la organización (Data Academy).

INDICADORES DEL PROYECTO

- Población cubierta por servicios básicos de sanidad digital del Servicio Nacional de Salud Población
- Población beneficiada por servicios digitales de gestión patologías crónicas u otros servicios específicos en al área de e-salud, envejecimiento activo y saludable
- Número de procedimientos teletramitables que se crean o mejoran con la operación
- Número de personas impactadas por actuaciones de difusión, concienciación y dinamización encaminadas a favorecer el uso y demanda de redes, productos y servicios TIC
- Población cubierta por los servicios de salud