



Plataforma de Donación de Datos Médicos

Idea avanzada desde la Vicepresidencia de Investigación del CSIC

Propuesta técnica del Grupo de Computación Avanzada y e-Ciencia (IFCA)

Planteamiento

- Las aplicaciones de diagnóstico y tratamiento de un paciente en Medicina en el siglo XXI se benefician enormemente del uso de datos proporcionados por otros pacientes, que han experimentado en muchos casos la misma situación
- Los datos disponibles sobre un paciente en su historia clínica detallan hasta tal punto su perfil, que es muy difícil en la práctica lograr su anonimización. Estos datos además contienen información relevante no solo para la situación tratada, sino para otros muchos aspectos relevantes de su vida personal, y ello se extiende a su familia, especialmente en el caso de datos genéticos.
- Pensamos que al facilitar el acceso a estos datos, la situación en cierto modo es similar a la de una donación de órganos: el paciente debe conocer la importancia de la donación, y el impacto de la misma.
- Por ello, proponemos que **se articule un protocolo de donación de datos médicos** que siga una estructura similar a la existente ahora mismo en Sanidad para la donación de órganos, y cuyo control resida íntegramente en manos públicas.
- Pero esta donación de datos requiere de una plataforma específica para su explotación, y aprovechamiento desde el sistema de sanidad

Desafíos técnicos

La Plataforma plantea múltiples retos técnicos:

-la definición de zonas de seguridad (“data safe haven*”) de depósito de los datos, que deben integrar adecuadamente recursos de computación que permitan explotar esos datos (típicamente mediante herramientas de Ciencia de Datos, entre las que destacan las conocidas como de *deep learning*)

-el desarrollo de software que garantice la trazabilidad en el uso posterior de los datos, y la transparencia del producto final desarrollado que hace uso de los mismos, evitando un uso discriminatorio de los mismos.

-la implementación de un derecho de retracción de los pacientes y sus familias, cumplido un límite temporal, sin perjudicar los avances logrados.

-la integración de diferentes zonas de seguridad de forma que se pueda realizar un estudio global sin necesidad de desplazamiento de los datos desde los diferentes sitios (típicamente hospitales) : un problema de computación distribuida.

* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4595892/pdf/btv279.pdf>

Desafíos temáticos (médicos)

La Plataforma plantea igualmente múltiples retos temáticos:

- la gestión específica de datos médicos, para asegurar su carácter “FAIR” (formatos, metadatos, ontologías)
- las herramientas de curación, y especialmente de validación, por profesionales médicos de los datos de partida
- la conexión con repositorios externos relevantes (incluyendo una posible descarga e integración)
- la definición de los métodos y herramientas de validación de las aplicaciones desarrolladas
- la discusión de los protocolos para abordar un posible uso clínico

Piloto propuesto

Se propone un piloto inicial de plataforma (desde el punto de vista técnico) que estaría en la línea de lo que se puede desarrollar posteriormente:

- Implementación de un “*deep* safe haven” en un centro existente de procesado de datos con recursos ya existentes suficientes para realizar las pruebas necesarias (la propuesta sería el CPD del Instituto de Física de Cantabria, CSIC-UC)

- Elección de un caso de uso para el que exista la posibilidad de contar con datos en formato realista, de los integrados en la historia clínica, preferiblemente anonimizado, y en principio pueda beneficiarse de un análisis que requiera un uso computacional intensivo (la propuesta inicial vendría a través del IDIVAL – HUMV)

En principio se accedería a dichos datos mediante la firma de un MTA

- Definición de soluciones desde el punto de vista médico (IDIVAL), biomédico (IBBTEC-CSIC-UC) y de Ciencia de Datos (IFCA-CSIC-UC) e implementación de las mismas (mediante COMPRA PUBLICA INNOVADORA, empresas TIC).

- Desarrollo en paralelo de un protocolo de donación de datos (Departamentos legales de IDIVAL y de Sanidad) que permita su automatización

Beneficios

- Desarrollo de una primera generación de aplicaciones específicas, cuya utilidad potencial quede demostrada en el piloto
- Análisis de problemas y soluciones encontrados en el proyecto, de cara a su escalado a nuevos casos de uso y a su implementación final en hospitales o en centros conectados a la red del Instituto de Salud Carlos III
- Impacto a corto/medio plazo en la atención al paciente, mediante la implementación de herramientas relativamente sencillas (“asistentes”) que proporcionen información de diagnóstico (desde filtrado a recomendaciones orientativas) en base a la historia clínica del paciente.
- Consolidación del prestigio y excelencia del HUMV, gracias al IDIVAL
- Posible propuesta de nuevas áreas de negocios en las empresas implicadas o generación de nuevas empresas de base tecnológica (EBT) de desarrollo de este tipo de soluciones

Plataforma para el Análisis de Datos Médicos

David Rodríguez González
Instituto de Física de Cantabria

Motivación

- Aplicar los avances en ciencia de datos para la investigación en el área de la salud utilizando datos clínicos, biológicos, sociales y medioambientales.
- Reusar datos ya existentes para investigación, bien sean datos clínicos, o datos de investigación agregados y curados que ayuden a formar un conjunto de datos suficientemente grande como para responder a preguntas de investigación con la potencia estadística necesaria.
- Ayudar a desarrollar un sistema de acceso a los datos de investigación que sea ágil, pero que a la vez mantenga estándares de calidad y gobernanza de la información.
- A largo plazo la ambición sería contribuir al avance de la medicina personalizada y tener un impacto en la salud de la población y en la mejora de los sistemas sanitarios.

Objetivos

- Facilitar el uso de herramientas de ciencia de datos (como el aprendizaje automático) con datos médicos.
- Desarrollar herramientas de ayuda al diagnóstico.
- Proporcionar un entorno seguro para el análisis y reutilización de datos médicos con:
 - Mecanismos bien establecidos de gobernanza de la información:
 - Incluido desidentificación.
 - Controles de acceso.
 - Auditoría.

Fuentes de datos

- Donación de datos
- Datos rutinarios
 - Mediante un proyecto y con los necesarios permisos éticos
- Depósito de datasets de investigación
 - Mediante acuerdo con el grupo de investigación
 - Temporal
 - permanente
 - ¿Costes?

Donación de datos

- Programa voluntario mediante el cual los usuarios del SCS puedan “donar” sus datos para su uso en investigación biomédica.
- Los datos serán “desidentificados”, no anonimizados para el depósito.
- Los usuarios podrán retirar su consentimiento en cualquier momento y los datos dejarán de utilizarse a partir de ese momento.
 - Por supuesto no se pueden retirar los datos de estudios pasados o cuando hayan sido agregados o anonimizado dejando de ser datos personales.
- Ejemplo en Escocia: SHARE <https://www.registerforshare.org/>



[MEET THE TEAM](#)

[RESEARCHERS](#)

[PARTNERS](#)

[CONTACT US](#)

[SHARE](#)

[STUDIES](#)

[WHY I SHARED](#)

[REGISTER](#)

[PRIVACY & SECURITY](#)

[WHAT'S NEW](#)

[QUESTIONS?](#)

[SPARE BLOOD](#)

[GET INVOLVED](#)



SAVE LIVES

Sign up now.

It takes less than a minute to submit your details

REGISTER

TELL YOUR FRIENDS

Like 356

Tweet

SHARE YOUR STORY

Data Safe Haven

- Infraestructura que permite almacenar, procesar y analizar datos personales.
- “Walled garden approach”
 - Los datos se almacenan, gestionan y analizan dentro del DSH.
 - No se pueden extraer los datos normalmente
 - podría considerarse la posibilidad en casos especiales
- Se pueden extraer resultados, siempre que no contengan datos personales.

Record Linkage Service

- Servicio que permite asociar los datos de un mismo individuo provenientes de diversas fuentes.
- Se puede mantener separado del DSH para casos en que los datos vengan ya anonimizados
 - Esto requeriría un sistema más complejo.

Plataforma de Análisis

- Asociada al DSH
 - Permite analizar los datos dentro del DSH
- Catálogo de aplicaciones
 - Podemos considerar algunas “estándar” para las diversas áreas
 - Pero el foco debe de estar en explotar herramientas de ciencia de datos
 - Deep learning en particular

Prototipo

- Aprovechar la experiencia y los equipos existentes en Cantabria:
 - SCS
 - UC
 - CSIC
- Selección de un número limitado (¿3?) de casos de uso que sirvan como ejemplo.
 - Imagen médica:
 - Clasificación de enfermedades psiquiátricas mediante MRI cerebral
 - Hemorragias cerebrales mediante TAC
 - Historial médico:
 - Clasificación de enfermedades hepáticas
 - Datos genéticos:
 - Pendiente (IBBTEC)

Prototipo

- Almacenamiento seguro
- Despliegue de aplicaciones a través de “contenedores verificados”
- Construir sobre proyectos actuales
 - H2020 DEEP
- Colaboración entre investigadores clínicos/biomédicos e investigadores en ciencia de datos.
- Protocolo de acceso a los datos para investigadores
 - Formación

Cantabria

- Población estable en torno a 600.000 personas.
- 3 núcleos de población urbanos destacables: Santander, Torrelavega, Laredo, y un alto componente de población dispersa
- Alto nivel de autosuficiencia desde el punto de vista sanitario

Hospital Universitario Marqués de Valdecilla

- Hospital de tercer nivel
- Acredita funciones de atención médica de alta calidad, investigación y docencia
- Soporte sobre el que se ha originado y desarrollado el Instituto de Investigación Valdecilla (IDIVAL) y el hospital Virtual Valdecilla
- 90 años de Historia (1929), altamente identificado y **valorado por la sociedad cántabra**
- Fomenta su marcada vocación histórica de hospital de referencia nacional para determinados servicios de alta cualificación asistencial y tecnológica

Experiencia en Donación en Cantabria

- Destacable, con alto grado de altruismo social desde la implantación del Sistema Público de Trasplantes de Órganos, y la Asociación de Donantes de Sangre.
- A la cabeza de España en la actividad trasplantadora por habitante: en un ranking internacional, cualquier comunidad autónoma española se encontraría en el primer puesto, ya que todas han batido récord al superar la tasa de 35 donantes por millón de personas. Nueve sobrepasan la media nacional (de 46,9) y cuatro superan incluso los 70, como Cantabria (74,1) La Rioja (71), País Vasco (70,3) y Navarra (70,3).
- A la cabeza de España en programas de donación: Cantabria actualmente tiene activos varios programas de donación altruista (sangre, órganos, tejidos, progenitores hematopoyéticos, leche), con unos índices de participación superiores en todos ellos a la media nacional, destacando especialmente el programa de donación de órganos y el de sangre.

Donación sangre /1000 hab.	2015	2016	2017	2018 (previsión)
España	36,6	36,6	36,3	36,6
Cantabria	40,12	38,34	37,6	39,06
Posición	6º	8º	8º	6º

- Donación de médula: donantes con tipaje HLA en REDMO (registro español donantes de MO) a 31 dic 2017: Cantabria 7,84 / 1000 hab. España 6,83 / 1000 hab.
- Donación de leche: éxito rotundo del primer año, superando expectativas (66 donantes).
- Donación de cordón: Desde hace 9 años también participamos en Programa nacional de donación de sangre de cordón umbilical. Nuestro índice es actualmente bajo.

- Originalmente el sistema de información fue diseñado (1991) e implementado en el Hospital Valdecilla, y posteriormente se ha extendido al resto de centros de la región. El equipo humano que lo ha desarrollado, funcionales y técnicos, sigue trabajando en su actualización y adecuación.
 - La aplicación “Visor corporativo de HCE” es la de mayor utilización por parte de los profesionales, en su práctica clínica diaria. Desarrollada y mantenida por HUMV. Permite acceder a todos los episodios asistenciales del paciente, originados indistintamente en Atención Primaria y Atención Especializada La información de pacientes mantiene una coherencia histórica desde su creación hasta la actualidad, con una alta calidad (ejemplo: grado de duplicidad de pacientes inferior a un 3x1000)
 - Trazabilidad en todos los contactos sanitarios regionales en pacientes, tanto en Atención Primaria como en Atención Especializada
 - Imagen Digital desde 2007, coherente y con alto grado de calidad del dato.
 - Desde el año 2002 la información de laboratorio se encuentra digitalizada, y estructurada a partir de 2012.
- Infraestructuras TIC renovadas
 - nuevo CPD, sistemas, comunicaciones, microinformática
- Garantizada la renovación de sistemas y su soporte en los próximos años

2017-PRUEBAS / TTOS MAS SOLICITADOS

Bioquímica (nº determinaciones)	6.734.689
Microbiología	396.834
Inmunología	195.650
Ecografías	59.248
Biopsias	40.288
Citologías	20.472
TAC	40.521
Determ, niveles sericos de farmacos	23.878
RMN	22.044
Mamografías	7.591
Gammagrafías	5.383
Radiología Intervencionista	8.202
Emg	5.193
Gastroscopias	4.427