

Ecosistema de datos sanitarios multipropósito para asistencia, gestión e investigación clínica

Miguel Pedrera-Jiménez

ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, España; Better, Liubliana, Eslovenia

Introducción

La historia clínica electrónica (HCE) se define, idealmente, como el repositorio longitudinal de datos sanitarios que se genera en torno al paciente, siendo su uso primario aquel que aplica a la asistencia del propio individuo¹. Adicionalmente, la HCE debe permitir su utilización para lo que se conoce como uso secundario, que abarca todo aquello adicional a la asistencia directa, como la gestión o la investigación clínica². En la práctica, la HCE es registrada y almacenada en diferentes sistemas de información sanitaria, con los que interactúan profesionales y pacientes para desplegar los procesos asistenciales y de uso secundario.

Sin embargo, estos sistemas de información siguen un diseño monolítico en el que cada aplicación y servicio implementa un modelo local para almacenar los datos necesarios para desplegar su funcionalidad³. Estas bases de datos locales no siguen estándares internacionales de modelado clínico⁴, sino que embeben la semántica de la información de salud en la propia aplicación. Esto lleva a una HCE fragmentada en múltiples silos de información, con una capacidad muy limitada de utilizar esos datos de manera combinada con significado pleno, e imposibilitando que el paciente esté realmente en el centro del ecosistema⁵.

Para resolver esta cuestión, en los últimos años ha surgido un nuevo paradigma de HCE que huye de la concepción monolítica de los sistemas de información y de HCE fragmentada, y que propone la construcción de una capa de datos estándar, neutral y común para el conjunto de los sistemas de información⁶. En

esencia, propone desligar la HCE de las aplicaciones, generando así datos centrados en el paciente y válidos para toda la vida.

El objetivo de este trabajo es proponer un marco técnico y organizativo para la construcción de un ecosistema de datos sanitarios multipropósito para asistencia, gestión e investigación clínica, trasladable a los procesos asistenciales y de uso secundario en el ámbito de la hematología.

Método

La metodología propuesta se basa en el uso de estándares abiertos de información sanitaria para la construcción de un ecosistema de datos sanitarios multipropósito para asistencia y propósitos de uso secundario. La validación de la metodología se llevó a cabo con la implantación de la plataforma INFOBANCO del Hospital Universitario 12 de Octubre, en Madrid, España⁷.

Paradigma de datos sanitarios multipropósito

Los datos sanitarios son empleados tanto para propósitos asistenciales, tales como la continuidad de la atención o el soporte a la toma de decisiones clínicas, como para propósitos secundarios, como la salud pública o la investigación clínica^{1,2}. Sin embargo, la fragmentación de la HCE en múltiples silos de información sanitaria impone un paradigma de «datos monodosis», en el que los datos se registran y validan múltiples veces para ser utilizados en un único propósito⁸.

Correspondencia:

Miguel Pedrera-Jiménez

E-mail: m.pedrera@alumnos.upm.es

Recibido: 06-07-2025

Aceptado: 06-08-2025

DOI: 10.24875/SANGRE.M25000061

Disponible en internet: 02-10-2025

Sangre. 2025;44(Supl 2):51-54

www.revistasangre.com

0036-4355 / © 2025 Sociedad Española de Hematología y Hemoterapia. Publicado por Permanyer. Éste es un artículo *open access* bajo la licencia CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

Esta metodología pretende alcanzar un nuevo paradigma de «datos sanitarios multipropósito», en el que los datos son recogidos y validados una única vez, en una HCE de semántica armonizada, y pueden ser empleados en múltiples propósitos. Para ello, es fundamental que los sistemas de información evolucionen desde la concepción de silos con bases de datos locales hasta la de agentes de un ecosistema de datos estándares neutrales a las aplicaciones⁷.

Estándares de información sanitaria

Los estándares de información sanitaria son especificaciones de convergencia para resolver necesidades de una manera común entre diferentes proveedores de datos. Entre estas necesidades encontramos el modelado clínico y la persistencia de la HCE, la interoperabilidad semántica de la información y la reutilización en investigación clínica⁹. Su uso es fundamental para construir esa capa de datos neutral a las aplicaciones y los servicios.

En la actualidad, los tres estándares principales empleados en el diseño de las plataformas de datos sanitarios de nueva generación son openEHR, HL7 FHIR y OMOP CDM. Los estudios recientes han establecido un marco de uso coherente de estos tres estándares¹⁰. Así, en la metodología propuesta, openEHR es el estándar adoptado para modelar y almacenar la HCE con pleno significado y contexto, HL7 FHIR para la interoperabilidad semántica y OMOP CDM para la reutilización en investigación con datos del mundo real.

Marco colaborativo de gestión de los datos sanitarios

La obtención de datos sanitarios multipropósito no solo pasa por adoptar estándares de información sanitaria, sino que es necesario definir un nuevo marco de trabajo colaborativo¹¹. Este acuerdo es fundamental si queremos una verdadera transformación digital en la organización, empoderando a los profesionales encargados del registro y la utilización de los datos.

Así, para los usuarios del ecosistema de datos multipropósito, se definieron un conjunto de obligaciones y derechos, recogido en la [tabla 1](#).

Plataforma INFOBANCO del Hospital 12 de Octubre

La plataforma INFOBANCO, diseñada e implementada por la Unidad de Ciencia de Datos del Hospital Universitario 12 de Octubre, constituye un demostrador

de ecosistema de datos sanitarios multipropósito⁹. En particular, la plataforma permite registrar, almacenar y acceder a datos para ofrecer los siguientes servicios a los profesionales de la organización:

- Servicios para asistencia: formularios y visores clínicos avanzados, listados dinámicos de cohortes, mecanismos de interoperabilidad y federación, y herramientas de vías clínicas compartidas y de soporte a la toma de decisiones.
- Servicios para gestión clínica: tablas dinámicas, cuadros de mando, mecanismos de interoperabilidad, federación y reutilización, y herramientas de evaluación y comparación (*benchmarking*) de resultados en salud.
- Servicios para investigación clínica: formularios y visores clínicos avanzados, listados dinámicos de cohortes, mecanismos de interoperabilidad, federación y reutilización, y herramientas analíticas de datos del mundo real.
- Servicios para desarrollo y evaluación de tecnología: entorno *sandbox* de datos sanitarios para diseño e implementación de herramientas asistenciales, de gestión y de investigación clínica.

Para ello, se definió una arquitectura de componentes basada en el uso agnóstico de los estándares de información de salud⁹. Así, se empleó la plataforma Better para el modelado y la persistencia armonizada, de acuerdo con openEHR, de la información procedente de una HCE monolítica¹², y desde este componente fue transformada y transferida a componentes basados en HL7 FHIR y OMOP CDM.

El flujo de datos entre los diferentes componentes de la plataforma INFOBANCO se muestra en la [figura 1](#).

Resultados

La plataforma INFOBANCO permitió combinar, de acuerdo con modelos de información comunes y estándares, datos de más de dos millones de pacientes a partir de los sistemas de información del Hospital Universitario 12 de Octubre y de atención primaria. Sobre esta capa de datos se desplegaron los diferentes servicios para asistencia, gestión, investigación y desarrollo de tecnología sanitaria.

Su primera aplicación fue durante la pandemia de COVID-19, permitiendo al hospital mejorar la asistencia y gestión de los pacientes, y participar en consorcios como ISARIC y 4CE¹³⁻¹⁵. Esta metodología fue considerada por los Centers for Disease Control and Prevention de los Estados Unidos de América como la metodología a seguir para la obtención y el intercambio de datos para investigación y salud pública¹⁶.

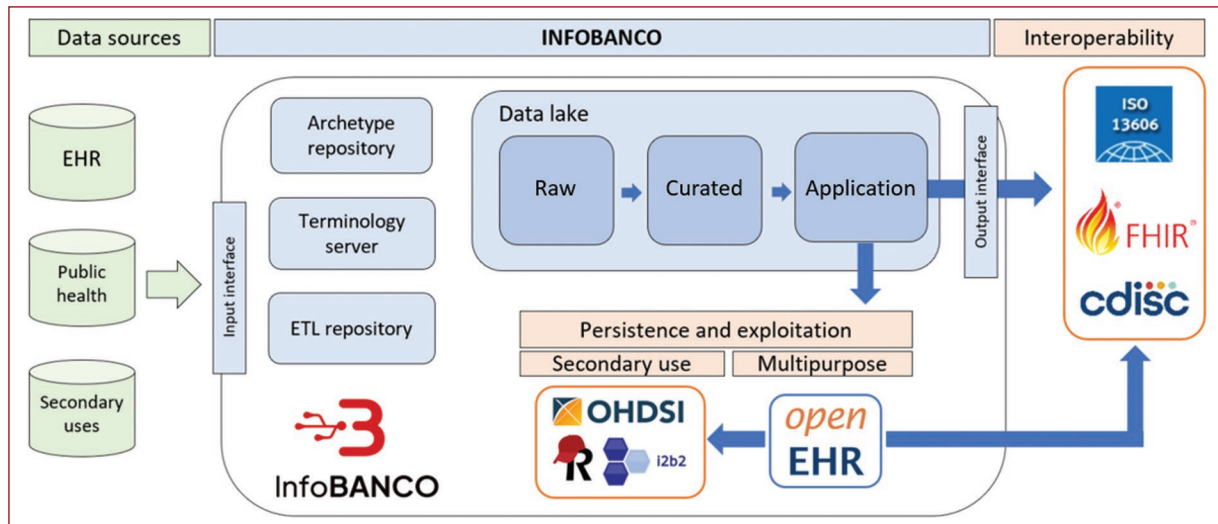


Figura 1. Esquema del flujo de datos entre los componentes de la plataforma INFOBANCO.

Tabla 1. Derechos y obligaciones de los usuarios del ecosistema de datos sanitarios multipropósito

Obligaciones	Derechos
O1. Los datos se registran dentro del ecosistema de datos multipropósito, evitando la generación de silos	D1. Acceso a un ecosistema de datos con la utilidad, la flexibilidad y la agilidad como principios de diseño
O2. El registro de datos es, dentro de lo posible, estructurado y basado en estándares	D2. Acceso seguro y controlado a los datos, respaldado por la tecnología adecuada
O3. La consulta, el intercambio y la explotación de los datos se realizan con las herramientas ofrecidas en el ecosistema	D3. Acceso a herramientas validadas de datos sanitarios para asistencia, gestión e investigación clínica

Posteriormente, estas herramientas fueron utilizadas por el servicio de hematología del Hospital Universitario 12 de Octubre. Así, se elaboraron listados dinámicos de pacientes y cuadros de mando clínicos para cada proceso asistencial definido en la cartera del servicio. Del mismo modo, a través de las herramientas analíticas de investigación con datos del mundo real, se publicaron diferentes artículos científicos en revistas de primer nivel del ámbito de la hematología^{17,18}.

Conclusiones

Este trabajo ha demostrado que la definición y la implementación de una HCE neutral a las aplicaciones

y servicios, y modelada, almacenada e intercambiada de acuerdo con los estándares internacionales, permite su utilización en múltiples propósitos sin necesidad de bases de datos de «usar y tirar». La plataforma INFOBANCO del Hospital Universitario 12 de Octubre es un demostrador de este nuevo paradigma de ecosistemas de datos sanitarios multipropósito para la asistencia, la gestión y la investigación.

Referencias

- Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform.* 2008;77:291-304.
- Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health. *J Am Med Inform Assoc.* 2007;14:1-9.
- Li E, Lounsbury O, Hasnain M, Ashrafian H, Darzi A, Neves AL, et al. Physician experiences of electronic health record interoperability and its practical impact on care delivery in the English NHS: a cross-sectional survey study. *BMJ Open.* 2025;15:e096669.
- Beale T. Archetypes: constraint-based domain models for future-proof information systems. *OOPSLA 2002 Work Behav Semant.* 2001;1-69.
- Bernstam EV, Warner JL, Krauss JC, Ambinder E, Rubinstein WS, Komatsoulis G, et al. Quantitating and assessing interoperability between electronic health records. *J Am Med Inform Assoc.* 2022;29:753-60.
- Piera-Jiménez J, Carot-Sans G, Ramiro-Pareta M, Nogueras MM, Folguera-Profits J, Ródenas P, et al. A 25-year retrospective of health IT infrastructure building: the example of the Catalonia Region. *J Med Internet Res.* 2024;26:e58933.
- Pedrera-Jiménez M, García-Barrio N, Rubio-Mayo P, Tato-Gómez A, Cruz-Bermúdez JL, Bernal-Sobrino JL, et al. TransformEHRs: a flexible methodology for building transparent ETL processes for EHR reuse. *Methods Inf Med.* 2022;61(S 02):e89-e102.
- Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, Terriza-Torres AI, López-Jiménez EA, Calvo-Boyero F, et al. Obtaining EHR-derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models. *J Biomed Inform.* 2021;115:103697.
- Pedrera-Jiménez M, García-Barrio N, Frid S, Moner D, Boscá-Tomás D, Lozano-Rubí R, et al. Can OpenEHR, ISO 13606, and HL7 FHIR work together? An agnostic approach for the selection and application of electronic health record standards to the next-generation health data spaces. *J Med Internet Res.* 2023;25:e48702.

10. Tsafnat G, Dunscombe R, Gabriel D, Grieve G, Reich C. Converge or collide? Making sense of a plethora of open data standards in health care. *J Med Internet Res*. 2024;26:e55779.
11. Frigidis LL, Chatzoglou PD. Implementation of a nationwide electronic health record (EHR). *Int J Health Care Qual Assur*. 2018;31:116-30.
12. Plataforma Better. Disponible en: <https://www.better.care/es/plataforma-better/>.
13. Pedrera M, García N, Blanco A, Terriza AI, Cruz J, López EA, et al. Use of EHRs in a tertiary hospital during COVID-19 pandemic: a multi-purpose approach based on standards. *Stud Health Technol Inform*. 2021;281:28-32.
14. Pedrera-Jiménez M, García-Barrio N, Rubio-Mayo P, Maestro G, Lalueza A, García-Reyne A, et al. Making EHRs trustable: a quality analysis of EHR-derived datasets for COVID-19 research. *Stud Health Technol Inform*. 2022;294:164-8.
15. Pedrera-Jiménez M, García-Barrio N, Hernández-Ibarburu G, Baselga B, Blanco A, Calvo-Boyero F, et al. Building an i2b2-based population repository for COVID-19 research. *Stud Health Technol Inform*. 2022;294:287-91.
16. Michaels M, Syed S, Lober WB. Blueprint for aligned data exchange for research and public health. *J Am Med Inform Assoc*. 2021;28:2702-6.
17. Sanz A, Ayala R, Hernández G, Lopez N, Gil-Alos D, Gil R, et al. Outcomes and patterns of treatment in chronic myeloid leukemia, a global perspective based on a real-world data global network. *Blood Cancer J*. 2022;12:94.
18. López-Muñoz N, Hernández-Ibarburu G, Alonso R, Sanchez-Pina JM, Ayala R, Calbacho M, et al. Large-scale real-life analysis of survival and usage of therapies in multiple myeloma. *J Hematol Oncol*. 2023;16(1):76. [published correction appears in *J Hematol Oncol*. 2023;16:81].