

White Paper

Revolucionando la salud: el potencial del procesamiento de lenguaje natural para la evaluación

CARLES ILLA, Senior Principal, Consulting Spain

INÉS JAUMOT MARCHANTE, Senior Consultant, Thought Leadership Spain

BORJA MARABINI, Engagement Manager, Consulting Services

CRISTINA DE LA TORRE, Consultant, Consulting Spain



Tabla de contenidos

Introducción	1
Aplicación de PLN en historias clínicas en el Servei de Salut de les Illes Balears	2
Aprendizajes del proceso de implementación de la nueva tecnología	4
Interés y expectativas del procesamiento de lenguaje natural	6
Conclusiones	7
Agradecimientos	7
Anexo	8
Bibliografía	9

Introducción

La Historia Clínica Digital del Sistema Nacional de Salud (HCDSNS) constituye una herramienta clave para la interoperabilidad de la información sanitaria en España. Este sistema permite a las Comunidades Autónomas compartir datos clínicos relevantes, garantizando su disponibilidad en formato electrónico en cualquier servicio autonómico del Sistema Nacional de Salud (SNS)¹. Mientras que las Comunidades Autónomas son responsables de implementar los requisitos técnicos y funcionales necesarios para la integración con el servicio HCDSNS, el Ministerio de Sanidad actúa como coordinador y facilitador del proceso. Actualmente, la cobertura poblacional de la HCDSNS alcanza un promedio del 91% en España², superando el 85% en la mayoría de las Comunidades Autónomas, con la excepción de Aragón, que presenta una cobertura del 52%³.

En el ámbito europeo, el *EU-Patient Summary* (Resumen de Paciente de la Unión Europea)⁴ permite a los profesionales sanitarios de distintos países acceder a la información clínica más relevante de un paciente que requiera atención no programada fuera de su país de origen. Este documento se genera a partir de los datos contenidos en la historia clínica electrónica del SNS, vinculados a la tarjeta sanitaria activa del paciente. Varias Comunidades Autónomas —como Madrid, Cataluña o la Comunidad Valenciana— y países europeos —como Portugal, Francia o Grecia— ya están integrados en este sistema⁵. Además, la iniciativa del Espacio Europeo de Datos Sanitarios (EEDS)⁶ busca ampliar este acceso, permitiendo el uso de datos sanitarios en toda Europa.

La historia clínica electrónica no solo ha mejorado la continuidad asistencial, sino que también ha abierto nuevas posibilidades para la explotación de datos con fines clínicos, de gestión e investigación⁷. No obstante, uno de los principales desafíos radica en que una gran parte de la información médica se registra en formato de

texto libre. Se estima que este tipo de datos representa hasta el 80% del total de la información médica^{8,9}. En consecuencia, el uso de técnicas de procesamiento de lenguaje natural (PLN) se ha vuelto esencial para extraer conocimiento útil a partir del análisis de texto libre^{10,11}, como lo demuestra la creciente literatura científica que apalanca esta tecnología.

En este contexto, IQVIA y el *Servei de Salut de les Illes Balears* han colaborado en un proyecto orientado a identificar intervenciones de dudoso valor o potencialmente inadecuadas, mediante el análisis de datos mayoritariamente no estructurados. Este proyecto, alineado con los principios del enfoque *Right Care*¹⁰, ha demostrado cómo el uso de tecnologías avanzadas de procesamiento de datos puede ser determinante para el éxito. Con el objetivo de profundizar en los beneficios, factores críticos de éxito y barreras para la implementación de este tipo de iniciativas, se han realizado entrevistas a diversos profesionales del *Servei de Salut de les Illes Balears*, incluyendo perfiles técnicos y de gestión, que participaron activamente en el proyecto.



Aplicación de PLN en historias clínicas en el *Servei de Salut de les Illes Balears*

Como se mencionó en el apartado anterior, uno de los principales desafíos para la generación de indicadores útiles en la gestión clínica es la disponibilidad de datos estructurados y fácilmente analizables. Sin embargo, una gran parte de la información clínica se encuentra registrada en formato de texto libre, lo que requiere un tratamiento previo antes de poder ser explotada, constituyendo así una barrera significativa para la obtención de indicadores de calidad.

En el marco de este proyecto, se priorizaron tres intervenciones de dudoso valor identificadas a partir del Compromiso por la Calidad de las Sociedades Científicas, en su vertiente de Atención Primaria. Este

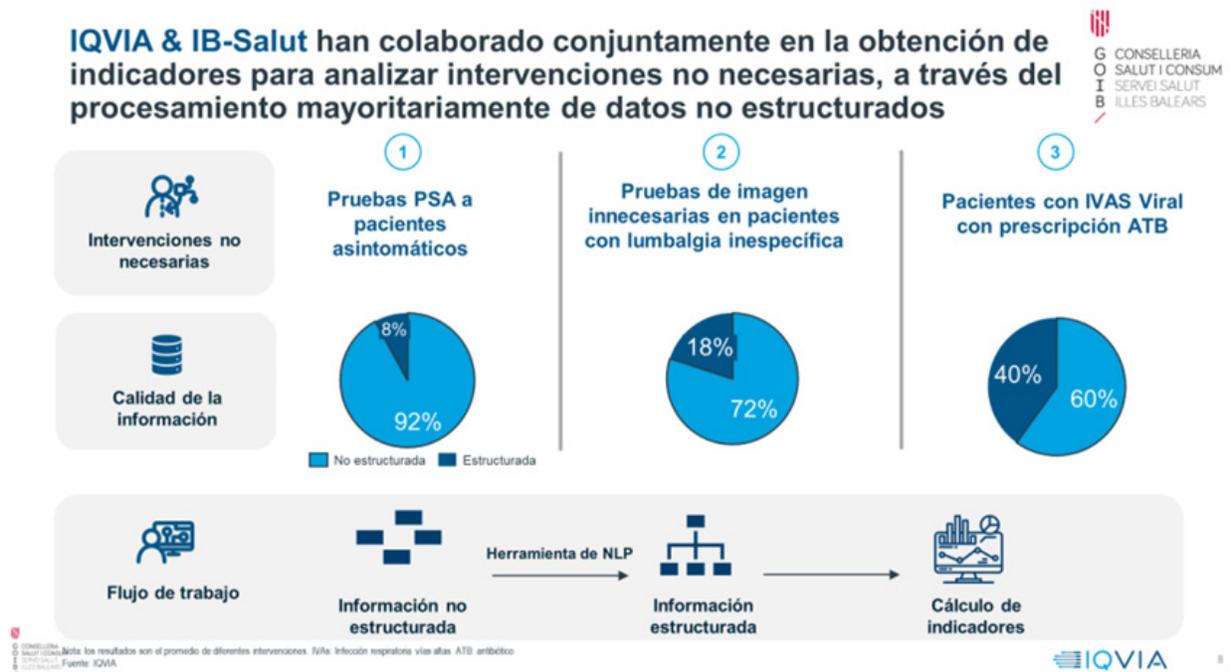
compromiso incluye las denominadas recomendaciones de “*No hacer*”, que consisten en prácticas clínicas que, según la evidencia científica, no aportan beneficios significativos al paciente y, en algunos casos, pueden incluso resultar perjudiciales o generar un uso ineficiente de los recursos sanitarios. Como ocurre con muchas de estas recomendaciones, la medición de las intervenciones dependía en gran medida de información no estructurada. En concreto, el porcentaje de datos no estructurados necesarios para el análisis fue del 92%, 72% y 60% respectivamente para cada una de las intervenciones seleccionadas (ver Figura 1). Esto significa que, en el mejor de los casos, solo se disponía de un 40% de información estructurada lista para su análisis.



Para superar esta limitación, se aplicó una herramienta de procesamiento de lenguaje natural (PLN) que permitió transformar el texto libre en datos estructurados, posibilitando así el cálculo de los indicadores deseados. Este elevado volumen de información no estructurada no es un caso aislado: por ejemplo, en Galicia se estima

que se generan unas 200.000 notas clínicas diarias¹¹. En el ámbito privado, el grupo hospitalario Ribera Salud identificó que el 40% de la información clínica estaba mal registrada, lo que refuerza la magnitud del problema¹².

Figura 1. Intervenciones de dudoso valor analizadas



Los profesionales entrevistados coinciden en que, en Baleares, los datos de atención primaria están significativamente menos estructurados que los datos hospitalarios. En particular, los antecedentes y diagnósticos suelen registrarse como texto libre, a pesar de existir campos específicos para su codificación. Entre los factores que contribuyen a esta falta de estructuración se mencionan: la ausencia de controles de calidad en la recogida de datos, la falta de revisión posterior, el desconocimiento o escasa formación del personal médico en el uso del *software*, la mayor comodidad de escribir en texto libre, y especialmente, la falta de un retorno tangible para el clínico por el tiempo adicional que implica registrar los datos de forma estructurada.

Finalmente, todos los entrevistados coincidieron en que este tipo de proyectos ha permitido generar indicadores valiosos y formular recomendaciones bien acogidas por los equipos participantes. Aunque muchas de estas recomendaciones eran previsibles, los indicadores aportaron una cuantificación objetiva de las intervenciones a eliminar. Desde una perspectiva técnica, el proyecto también demostró que es posible trabajar directamente con datos no estructurados, evitando así la necesidad de invertir grandes esfuerzos en su transformación, estandarización o normalización previa.

Aprendizajes en la implementación de tecnologías de PNL en el ámbito sanitario

La implementación de proyectos innovadores en el sector sanitario, especialmente aquellos relacionados con tecnologías digitales, mejora de la calidad o eficiencia, presenta una alta tasa de fracaso. Algunas referencias internacionales sitúan esta tasa entre el 30% y el 90%¹³. Diversos estudios señalan que el éxito de estas iniciativas depende de múltiples factores, como el tiempo de implementación, la forma de comunicación y el perfil de los receptores de la innovación^{14,15}. En este sentido, la teoría de la Difusión de Innovaciones¹⁶ identifica cinco características clave que influyen en la adopción temprana: ventaja relativa, compatibilidad, complejidad, posibilidad de prueba y visibilidad del impacto.

En el proyecto analizado, los entrevistados identificaron barreras comunes en la implementación de iniciativas innovadoras¹⁷, como la gestión del cambio, la sostenibilidad financiera y la necesidad de nuevos perfiles profesionales con capacidades específicas. Estas barreras afectaron al proyecto en distintas fases: diseño, ejecución y continuidad.

Fase de diseño

Una de las principales áreas de mejora señaladas fue el desconocimiento inicial sobre la información no estructurada disponible. Al no saber con certeza qué tipo de datos contenían los campos de texto libre, existía el riesgo de no obtener información relevante para los objetivos del proyecto. Por ello, los profesionales entrevistados sugieren realizar un análisis preliminar de los datos disponibles, como cuantificar cuántos profesionales completan los campos MEAPs (Motivo de consulta, Examen clínico, Análisis y orientación diagnóstica, Plan y tratamiento), para estimar el volumen potencial de información útil.

Asimismo, se destacó la importancia de garantizar la seguridad de la información desde el inicio. La falta de conocimiento del proyecto por parte de los equipos responsables de los datos generó reticencias, que se resolvieron mediante el diseño de mecanismos que aseguraran el cumplimiento del Esquema Nacional de Seguridad (ENS).

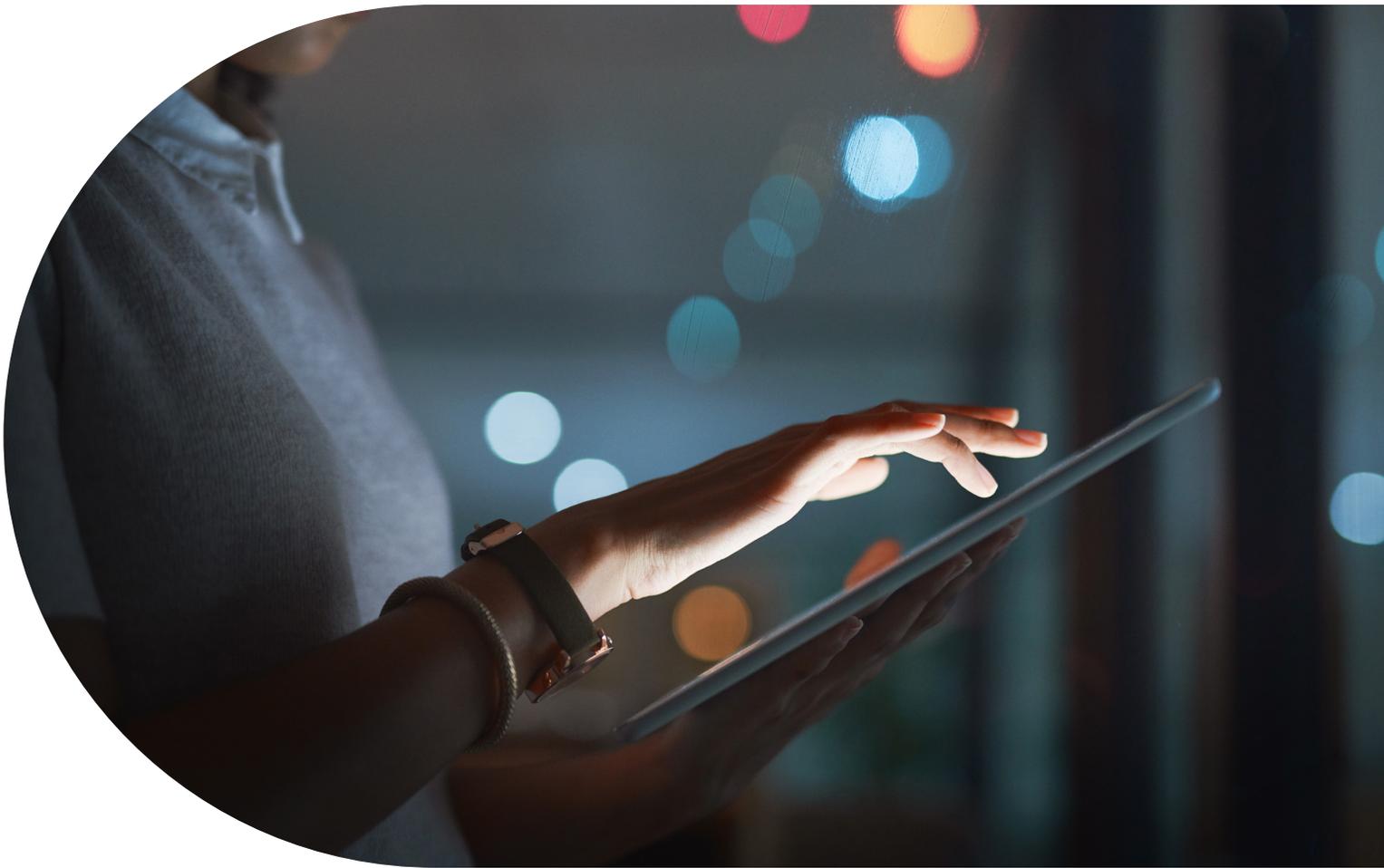
Fase de explotación

Aunque no afectó directamente a este proyecto, se identificó como potencial barrera la falta de consenso en el uso de escalas clínicas. La existencia de múltiples metodologías para el seguimiento de patologías otorga flexibilidad al profesional, pero dificulta la explotación homogénea de los datos.

Fase de ejecución

Durante la ejecución, una de las principales dificultades fue la gestión de tareas y prioridades. La participación de personal interno del sistema de salud, aunque bien intencionada, se vio limitada por la carga de trabajo habitual. Las tareas del proyecto quedaban relegadas frente a otras más urgentes, y las incidencias técnicas se resolvían con lentitud por falta de personal dedicado.

Además, los plazos cerrados impuestos por la financiación dificultaron una implementación gradual y una adecuada gestión del cambio. Esta presión temporal impidió una mayor implicación de perfiles asistenciales, una personalización de los resultados y una mayor visibilidad del proyecto. Aunque los resultados se pusieron a disposición de todos los usuarios, no se les ha dado un uso activo. Los entrevistados señalaron que este tipo de iniciativas tecnológicas se perciben como una “caja negra” por parte de los clínicos, generando desconfianza sobre su funcionamiento. Habrían preferido una mayor implicación de los profesionales en el diseño y desarrollo del proyecto, fomentando así la co-creación y la apropiación de los resultados.



Como aprendizaje, en un proyecto posterior se asignó una persona responsable de comunicar los hitos alcanzados, lo que mejoró la visibilidad y el compromiso.

Para escalar y dar continuidad al proyecto, se identificaron dos elementos clave:

- Nuevas capacidades profesionales: La implementación y mantenimiento de tecnologías innovadoras requiere perfiles técnicos específicos. Actualmente, el sistema sanitario depende en gran medida de financiación y soporte externos, lo que compromete la sostenibilidad una vez finalizado el proyecto.
- Involucración de *stakeholders*: Se considera fundamental implicar a más perfiles directivos

desde el inicio, para asegurar que todos los departamentos potencialmente implicados estén alineados y comprometidos. Esto facilitaría la continuidad del proyecto incluso si cambia el equipo responsable.

En resumen, existe consenso entre los profesionales entrevistados sobre la necesidad de impulsar iniciativas innovadoras en el ámbito sanitario. Sin embargo, también reconocen la dificultad de integrar nuevas formas de trabajo en la práctica diaria. Esta resistencia al cambio, inherente a las innovaciones disruptivas, limita que los propios profesionales propongan nuevas formas de trabajar, debido al desconocimiento tanto de las tecnologías disponibles como de su potencial aplicación en el entorno asistencial.

Interés y expectativas futuras en el uso de PNL en salud

Los entrevistados señalaron que, dentro del *Servei de Salut de les Illes Balears*, ya existían grupos de trabajo centrados en la identificación de prácticas clínicas a eliminar (“*No hacer*”). En algunos casos, estas prácticas alcanzaban hasta un 80% de intervenciones que deberían evitarse. El interés por dejar de implementar estas actuaciones no se limita únicamente al ahorro económico que suponen, sino también al impacto positivo en la eficiencia del sistema: reducción del tiempo dedicado por los profesionales sanitarios, disminución de las listas de espera y, en consecuencia, beneficios tangibles para gestores, clínicos y pacientes.

Este proyecto ha demostrado el potencial del procesamiento de lenguaje natural (PLN) como herramienta eficaz para generar indicadores clave en tiempos muy reducidos. El PLN, ampliamente utilizado en investigación médica, ofrece un gran margen de desarrollo en el ámbito asistencial. Tal como se ha evidenciado en este estudio, permite calcular indicadores a partir de texto libre, pero también existen otras aplicaciones documentadas, como:

- Identificación de pacientes susceptibles de ser atendidos por telemedicina para reducir listas de espera

- Priorización de pacientes con comorbilidades¹⁷
- Codificación automática de diagnósticos^{18,19}
- Detección de efectos adversos²⁰
- Incorporación de reconocimiento de voz para facilitar la introducción de datos en la historia clínica²¹

A futuro, los entrevistados expresaron una visión optimista sobre la integración de herramientas de inteligencia artificial (IA) en la práctica clínica diaria, especialmente en atención primaria. Imaginan un escenario en el que una herramienta de PLN sea capaz de codificar automáticamente los problemas de salud y estructurar el texto libre que el médico escribe o dicta durante la consulta. A partir de ahí, una IA generativa podría sugerir el siguiente paso clínico, basándose en protocolos y recomendaciones del tipo “*Hacer, No hacer*”.

En el ámbito de la investigación, también se vislumbra un gran potencial para el PLN, ya que el texto libre aporta una riqueza de información que complementa y amplía los datos cuantitativos recogidos en campos estructurados. Asimismo, se mencionaron otros casos de uso emergentes, como la aplicación de IA en radiología y dermatología para apoyar el diagnóstico por imagen.

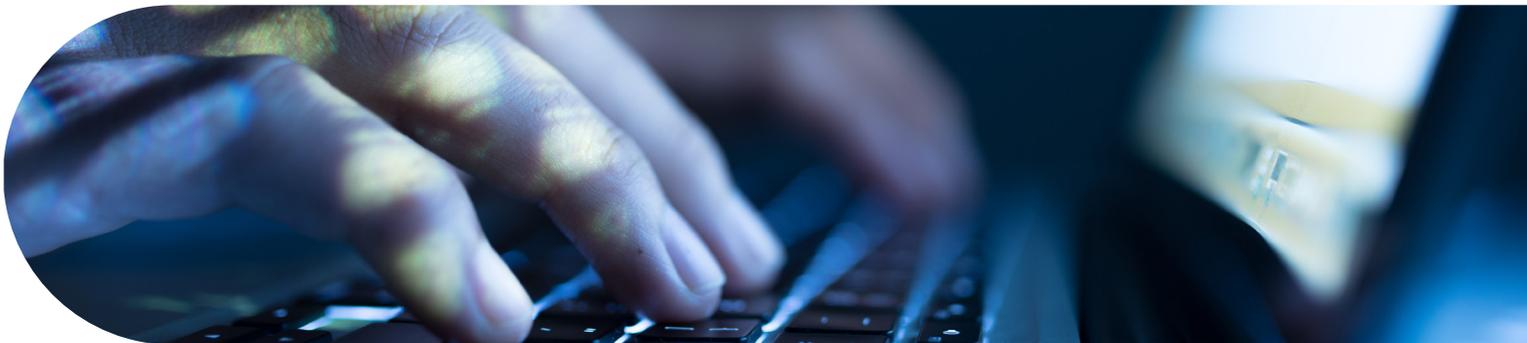
Conclusiones

La disponibilidad de datos clínicos analizables sigue siendo uno de los principales retos del sistema sanitario para generar indicadores que permitan mejorar la atención al paciente. A través de la implementación de este proyecto, se ha constatado el potencial del procesamiento de lenguaje natural (PLN) como herramienta eficaz para explotar los datos no estructurados presentes en las historias clínicas electrónicas.

Este trabajo demuestra que, si bien trabajar con datos no estructurados representa una dificultad añadida, no constituye un impedimento. Las herramientas de PLN permiten transformar esta información en datos

estructurados, facilitando su análisis y reduciendo significativamente el tiempo necesario para obtener resultados. Además, una vez implementadas, estas soluciones permiten automatizar procesos de forma recurrente, lo que incrementa su eficiencia y sostenibilidad en el tiempo.

En definitiva, la aplicación del PLN contribuye a maximizar el valor de los datos clínicos no estructurados, lo que redonda en una mejora de la calidad asistencial, en la eficiencia del sistema y en la toma de decisiones basada en evidencia.



Agradecimientos

Queremos expresar nuestro agradecimiento a las siguientes personas del *Servei de Salut Illes Balears* (IB-SALUT) por su valiosa colaboración en este artículo:

- **Antoni Joan Colombras i Salas**, *Sotsdirector de Gestió i Governança de la Dada*
- **Patricia Lorente-Montalvo**, especialista en Medicina Familiar y Comunitaria. *Servei de Salut Illes Balears* (IB-SALUT). Instituto de investigación Sanitaria Illes Balears (IdISBa)
- **Daniel Miró Pettican**, *Tècnic de l'Observatori de Dades Sanitàries del Servei de Salut de les Illes Balears*
- **Rosa Ortuño Muro**, médico de familia. *Sistemes d'informació clínica*.
- **Noemí Pérez García**, *Tècnic de l'Observatori de Dades Sanitàries del Servei de Salut de les Illes Balears*
- **Antònia Salvà Fiol**, *Cap de l'Observatori de Dades Sanitàries del Servei de Salut de les Illes Balears*

Anexo

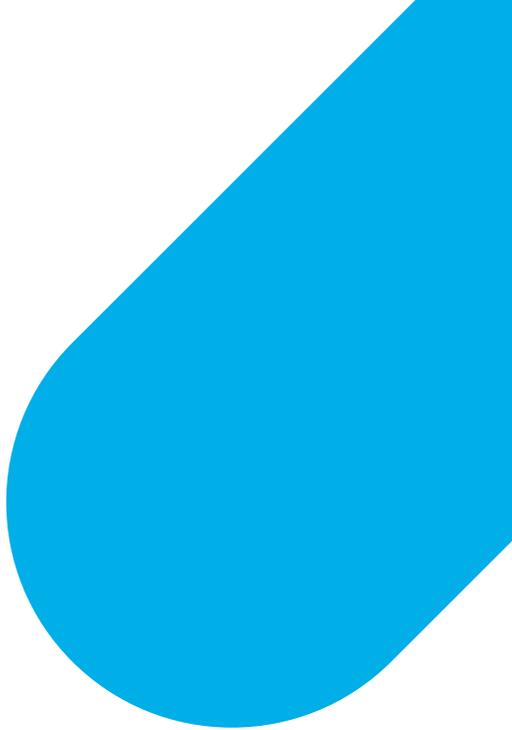
1. Tabla-resumen de los aprendizajes del proyecto

FASES DEL PROYECTO	BARRERAS Y DIFICULTADES DETECTADAS	RECOMENDACIONES Y POSIBLES ACCIONES
Diseño	Desconocimiento de la información desestructurada disponible	Realizar un análisis previo de los datos disponibles para asegurar que haya datos suficientes para sacar resultados.
Diseño	Seguridad de la información	Anticipar posibles impedimentos y planificar como gestionarlos para cumplir con la normativa vigente.
Diseño	Falta de consenso en las escalas	La utilización de distintas técnicas y escalas para el seguimiento de algunas patologías complica la explotación de los datos recogidos. Normalizar y estandarizar escalas ayudaría en este aspecto.
Ejecución	Gestión de tareas y prioridades	Contar con personal técnico principal o exclusivamente dedicado al proyecto ayudaría a resolver las incidencias más rápidamente.
Ejecución	Tiempo de gestión del cambio	Si no hubiera una presión temporal, se podría involucrar a más participantes (sobre todo roles asistenciales), conseguir un resultado más personalizado y ganar más visibilidad.
Ejecución y escalabilidad	Necesidad de roles con nuevas capacidades	Actualmente, el sistema sanitario suele requerir de fondos externos y de empresas externas para desarrollar iniciativas con nuevas tecnologías y poder darles continuidad.
Escalabilidad	Mayor involucración de <i>stakeholders</i>	Involucrar a más roles directivos para asegurar que todos los departamentos/servicios que puedan llegar a interactuar con este proyecto en algún momento (incluso post-implementación) se familiaricen e impliquen.
Escalabilidad y propuesta de nuevas iniciativas innovadoras	Resistencia al cambio	Es complejo adaptarse a los cambios en la forma de trabajar y que surjan nuevas propuestas por desconocimiento tanto de las nuevas tecnologías como de las oportunidades de integrarlas en el día a día.

Bibliografía

1. Historia Clínica del Sistema Nacional de Salud (HCDSNS). Ministerio de Sanidad. Disponible en: <https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaSNS/home.htm>
2. Situación actual de Historia Clínica Digital del SNS. Informes Clínicos disponibles en cada CCAA. Disponible en: <https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaSNS/mapa/situacionActualHCDSNS.htm>
3. Situación actual de Historia Clínica Digital del SNS. Informes Clínicos disponibles en cada CCAA. Actualización: Febrero 2025. Ministerio de Sanidad. Disponible en: https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaSNS/mapa/Tabla_HCDSNS.pdf
4. Historia resumida europea (EU-Patient Summary) (EUPS). Ministerio de Sanidad. Disponible en: <https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaUE/home.htm>
5. Países con servicio EU-Patient Summary. CCAA de España que pueden comunicar informes EU-Patient Summary (EUPS) y Estados Miembros UE con los que se puede operar. Disponible en: <https://www.sanidad.gob.es/areas/saludDigital/historiaClinicaUE/paisesEUPS.htm>
6. Espacio Europeo de Datos Sanitarios (EEDS). Web oficial de la Unión Europea. Disponible en: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_es
7. R.M. Añel Rodríguez, I. García Alfaro, R. Bravo Toledo et al. Historia clínica y receta electrónica: riesgos y beneficios detectados desde su implantación. Diseño, despliegue y usos seguros. Atención Primaria 53 (2021). Disponible en: <https://doi.org/10.1016/j.aprim.2021.102220>
8. Murdoch TB, Detsky AS. The Inevitable Application of Big Data to Health Care. JAMA. 2013;309(13):1351–1352. Disponible en: <https://doi.org/10.1001/jama.2013.393>
9. Kong HJ. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res. 2019 Jan;25(1):1-2. Disponible en: <https://doi.org/10.4258/hir.2019.25.1.1>
10. Elshaug, A., et al. Levers for addressing medical underuse and overuse: achieving high-value health care. The Lancet, SERIES|RIGHT CARE| VOLUME 390, ISSUE 10090, P191-202, JULY 08, 2017. Disponible en: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(16\)32586-7/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(16)32586-7/fulltext)
11. Krallinger M. Centro Nacional de Investigaciones Oncológicas. Actuaciones de la Oficina Técnica de Sanidad. Infoday sobre Tecnologías del Lenguaje en Sanidad. Disponible en: https://plantl.digital.gob.es/tecnologias-lenguaje/comunicacion-formacion/eventos/eventosinfoday2018/Actuaciones%20de%20la%20Oficina%20T%C3%A9cnica%20de%20Sanidad/O_Martin_krallinger.pdf
12. Educa News. La inteligencia artificial halla patrones en las historias clínicas donde tú no eres capaz de verlos. Disponible en: <https://educa-med.online/ribera-salud-inteligencia-artificial/>

13. Jacobs SR, Weiner BJ, Reeve BB, Hofmann DA, Christian M, Weinberger M. Determining the predictors of innovation implementation in healthcare: a quantitative analysis of implementation effectiveness. *BMC Health Serv Res.* 2015;15(1):6.
14. Stephenson R, Phelps A, Colburn J. Diffusion of innovations and Program Implementation in Areas of Health Behavior/Education/Promotion, Physical Activity, and Physical Education. *The ICHPER-SD journal of research in health, physical education, recreation. Sport Dance.* 2018;10(1):3–11.
15. Drury DH, Farhoomand A. Innovation Diffusion and implementation. *Int J Innov Manag.* 1999;03(02):133–57.
16. Rogers EM, Singhal A, Quinlan MM. Diffusion of innovations. In: Stacks DW, Salwen MB, editors. *An integrated approach to communication theory and research.* 2nd ed. London; New York: Routledge; 2008. pp. 418–34
17. Báez P, Arancibia AP, Chaparro MI, Bucarey T, Núñez F, Dunstan J. Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile. *Revista Médica Clínica Las Condes.* 2022 November-December;33(6):576–82. Disponible en: <https://doi.org/10.1016/j.rmcl.2022.10.002>
18. Henry S., Wang Y., Shen F., Uzuner O. The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records. *J Am Med Inform Assoc.* 2020;27(10):1529–1537. Disponible en: <https://doi.org/10.1093/jamia/ocaa106>
19. Blanco A., Perez-de-Viñaspre O., Pérez A., Casillas A. Boosting ICD multi-label classification of health records with contextual embeddings and label-granularity. *Comput Methods Programs Biomed.* 2020;188:105264. Disponible en: <https://doi.org/10.1016/j.cmpb.2019.105264>
20. Yu D, Vydiswaran VGV. An Assessment of Mentions of Adverse Drug Events on Social Media With Natural Language Processing: Model Development and Analysis. *JMIR Med Inform.* 2022 Sep 28;10(9):e38140. Disponible en: <https://doi.org/10.2196/38140>
21. Vivancos P. El reconocimiento de voz para mejorar la adopción de la historia clínica electrónica. I+S: Revista de la Sociedad Española de Informática y Salud, N° 155 (Junio), 2023, pág. 57. Disponible en: <https://seis.es/is-155/>



CONTACT US
iqvia.com

