

Automation of Cost Recovery in the public health system of the Autonomous City of Buenos Aires

Guerrero Giménez, María Luz¹ ²[0009-0009-6290-5407] and Bardaui, Ariana¹ ³[0009-0000-5020-7331]

¹ Gerencia Operativa de Gestión de Información y Estadísticas de Salud, Ministerio de Salud, Gobierno de la Ciudad Autónoma de Buenos Aires

² Universidad Nacional de La Plata

³ Universidad Nacional de La Matanza

Abstract. In Argentina, the health system operates under a mixed and decentralized model, combining different levels of management and financing. Within this framework, the cost recovery process in the public health sector of Buenos Aires City (CABA), regulated by Law 5.622 and its amendments, aims to recover the costs of health services provided to patients with insurance coverage. Following the implementation of the Electronic Health Record (EHR) system and the progressive digitalization of clinical data, the Health Information and Statistics Management Office (GOGIES) launched a document automation project to improve compliance with administrative procedures. As the system became more complex and data volume increased, it was necessary to redesign document generation processes. The proposed solution focused on building workflows based on Directed Acyclic Graphs (DAGs), using open-source tools such as R and GO. This enabled greater computational efficiency, scalability of report generation, and timely availability of documents for billing purposes.

Keywords: Recovery, Automation, Digitalization.

Automatización del Recupero de Gastos del sistema público de salud de la Ciudad Autónoma de Buenos Aires

Resumen. En Argentina, el sistema de salud se encuentra organizado bajo un modelo mixto y descentralizado, en el que coexisten diferentes niveles de gestión y financiamiento. Su estructura busca garantizar el acceso a la atención sanitaria a través de tres subsistemas que operan de manera complementaria. En este marco, el recupero de gastos en el subsistema público la Ciudad Autónoma de Buenos Aires (CABA), regulado por la Ley 5.622 y sus modificatorias, busca recuperar los costos de la atención de pacientes con cobertura de salud que reciben servicios en la Red Integral de Cuidados Progresivos del subsector público. A partir de la implementación de la Historia Clínica Electrónica (HCE) y la progresiva digitalización de los datos clínicos, la Gerencia Operativa de Gestión de Información y Estadísticas de Salud (GOGIES) inició un proyecto de automatización documental para optimizar el cumplimiento de los procedimientos administrativos vigentes. Con la creciente complejidad del

sistema y el volumen de datos, se hizo necesario rediseñar los procesos de generación documental. La solución se enfocó en la construcción de flujos de trabajo basados en grafos acíclicos dirigidos (DAGs), utilizando herramientas de código abierto como R y GO, lo que permitió mejorar la eficiencia computacional, escalar la generación de reportes y garantizar su disponibilidad oportuna para la facturación.

Palabras clave: Recupero de gastos, Automatización, Digitalización.

1 Introducción

El recupero de gastos del subsistema de salud público de la Ciudad Autónoma de Buenos Aires (CABA), reglamentado por la Ley 5.622 y sus modificatorias, es un proceso mediante el cual se busca recuperar los recursos utilizados al atender a pacientes con cobertura de obras sociales, empresas de medicina prepaga o similares. Este circuito constituye una herramienta fundamental para la sostenibilidad financiera de las instituciones que lo integran. En contextos como el argentino, donde coexisten pacientes con y sin cobertura formal de salud, el cobro a obras sociales, prepagas o seguros por prestaciones realizadas a sus afiliados representa una fuente clave de financiamiento.

Durante muchos años, la implementación de este recupero dependió de procesos manuales, en los que la generación de documentación asistencial y administrativa recaía sobre el personal de salud y operadores técnicos. Esta modalidad no solo implicaba una elevada carga operativa, sino que además dificultaba la estandarización, generaba inconsistencias entre registros y demoraba el circuito de facturación. La ausencia de mecanismos digitales para integrar información clínica, administrativa y de cobertura también limitaba la trazabilidad del proceso y el monitoreo del rendimiento del sistema.

2 Antecedentes

La implementación de la Historia Clínica Electrónica (HCE) en el ámbito público de la Ciudad en 2012 representó un cambio de paradigma en la disponibilidad, accesibilidad y calidad de la información sanitaria. A través de la HCE, se digitalizaron progresivamente los registros clínicos, de diagnóstico, intervenciones, internaciones y prestaciones ambulatorias. Esta transformación posibilitó la centralización y estandarización de datos antes dispersos, lo cual abrió nuevas oportunidades para automatizar procesos que antes requerían validación manual.

El procedimiento administrativo se inicia cuando FACOEP recibe los Comprobantes de Recupero de Gastos (CRG) enviados por los efectores públicos. Para la conformación de dicha documentación los centros de atención deben adjuntar, entre otros elementos, la documentación respaldatoria del asiento electrónico de la prestación brindada o Detalle de las Prestaciones Hospitalarias (DPH).

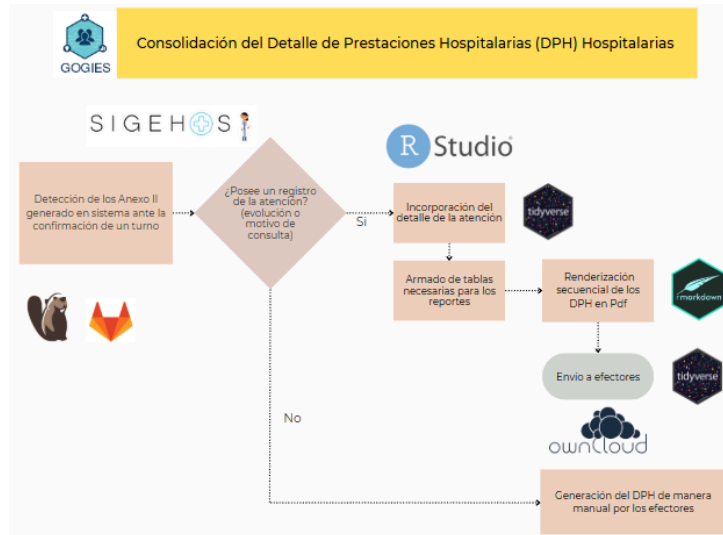


Fig. 1. Consolidación del Detalle de Prestaciones Hospitalarias (DPH) en R

Para cumplimentar este aspecto clave, la Gerencia Operativa de Gestión de Información y Estadísticas de Salud (GOGIES) desarrolló un proyecto para generar automáticamente el DPH con el detalle de las prestaciones realizadas por cada documento generado en sistema que certifican las atenciones ambulatorias y de guardias (Anexo II) para instancias asistenciales efectuadas con pacientes con cobertura. Este proceso caracteriza a los Anexos II con información de las atenciones a través de métodos basados en reglas vía expresiones regulares en texto clínico libre. Se trata de métodos interpretables y efectivos para la clasificación y extracción de información del texto médico (Dalianis, 2018).

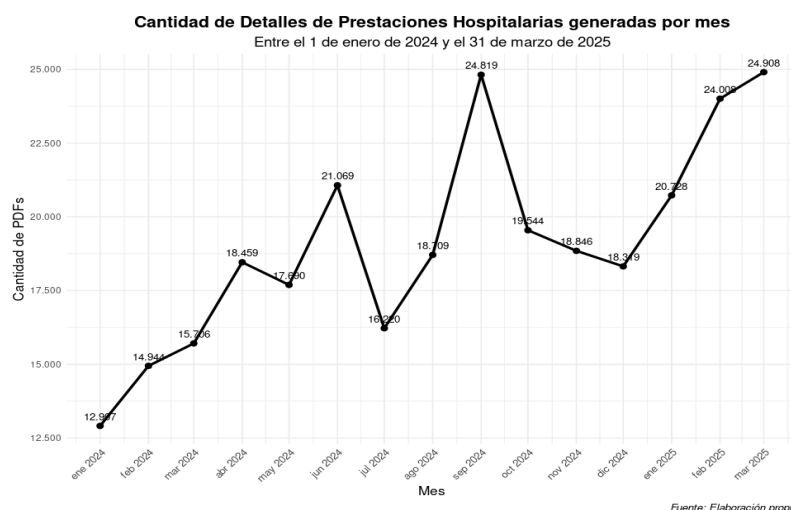


Fig. 2. Detalle de Prestaciones Hospitalarias (DPH) generadas entre el 1 de enero de 2024 y el 31 de marzo de 2025

No obstante, el constante avance en la digitalización trajo aparejados nuevos desafíos. Como se observa en la Fig.2, el crecimiento de los volúmenes de datos, la multiplicación de fuentes de información y la necesidad de garantizar la integridad de los registros demandaron soluciones más sofisticadas desde el punto de vista técnico.

El modelo desarrollado de generación documental basado en scripts desarrollados en R no podía escalar adecuadamente frente a estas nuevas exigencias. Procesos que antes podían resolverse con herramientas ofimáticas en un solo script comenzaron a presentar cuellos de botella: tiempos de procesamiento de más de 30 horas, errores de ejecución en entornos no controlados, dificultades para replicar tareas en distintos efectores y problemas de mantenimiento del código desarrollado.

Ante este escenario, la GOGIES identificó la necesidad de rediseñar el proceso de generación de documentación para el recupero de gastos, incorporando herramientas modernas de automatización y flujos de trabajo eficientes, escalables y auditables.

3 Solución propuesta

La solución adoptada consistió en el diseño e implementación de un sistema de automatización basado en flujos de trabajo estructurados como grafos acíclicos dirigidos (DAG - *Directed Acyclic Graph*), capaces de orquestar tareas interdependientes de forma modular. Un DAG es un grafo en el que todas las aristas están dirigidas. La ausencia de ciclos garantiza que, cuando se parte de cualquier nodo y se sigue la dirección es imposible volver al mismo nodo (Averin et al., 2025). Esta arquitectura permite que cada componente del flujo —como la lectura de datos, validación, integración, análisis y generación de reportes— se ejecute en etapas

diferenciadas, lo que facilita la paralelización, el control de errores y la trazabilidad de los procesos.

La decisión de adoptar DAGs como estructura base para el modelado del flujo de trabajo se sustenta en su capacidad para representar de forma modular y transparente dependencias entre componentes. Aunque su uso se ha popularizado en el ámbito de la salud pública para representar relaciones causales entre variables y seleccionar conjuntos de ajuste (Tennant et al., 2021), en este proyecto se emplearon como arquitectura computacional para organizar secuencias de tareas interdependientes.

Este tipo de abordajes resulta análogo al de plataformas como *Apache Airflow*, que también utilizan grafos acíclicos dirigidos para representar flujos de trabajo y sus dependencias internas (Shubha & Prasad, 2019). Este enfoque se alinea con prácticas consolidadas en las ciencias computacionales, donde los DAGs constituyen la base estructural de sistemas de orquestación de tareas, permitiendo ejecutar procesos complejos de manera escalable, eficiente y trazable. Estos sistemas contribuyen a mejorar el rendimiento general, la eficiencia operativa y la auditabilidad en entornos de procesamiento de grandes volúmenes de datos (Castejón-Limas et al., 2022).

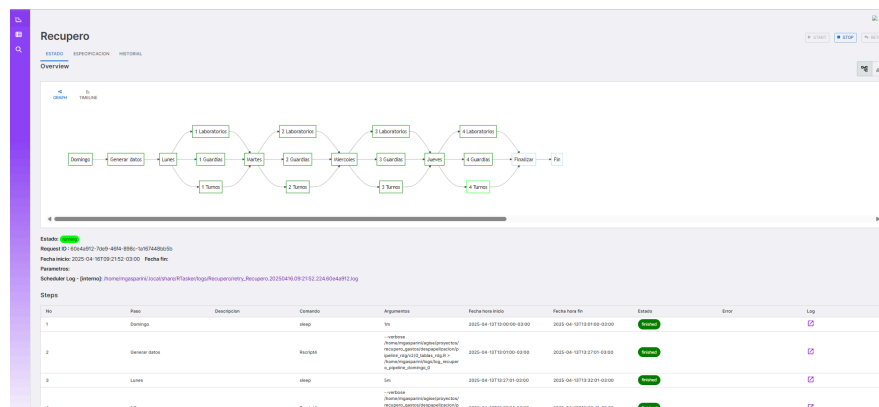


Fig. 3. Aplicativo generado con Diagrama de flujo acíclico dirigido (DAG) correspondiente al pipeline de generación y procesamiento semanal de datos.

3.1 Elección y diseño del flujo DAG

Con el fin de mejorar la eficiencia de un proceso que inicialmente requería más de 30 horas de ejecución, se dividió el script original en seis módulos funcionales. Esta compartimentación permitió descomponer el pipeline en tareas independientes y encadenadas, estructuradas como un DAG, donde cada nodo representa una tarea autónoma y cada arista una dependencia lógica entre procesos.

La segmentación se organizó, por un lado, en torno al consumo y modelado de las tablas necesarias para la construcción de los reportes; y por otro, según la fuente de

información y el día de procesamiento. Esto permitió ejecutar tareas en paralelo cuando no existían dependencias, reduciendo significativamente los tiempos de cómputo. Asimismo, se generaron reglas dinámicas para ejecutar el flujo de manera semanal, con actualización automática de los parámetros de fecha utilizados en la extracción y validación de datos.

3.2 Implementación técnica y monitoreo

Se optó por el uso de herramientas de código abierto, destacándose GO para el desarrollo de un aplicativo que permite la ejecución secuencial del procesamiento con mayor rendimiento, y R por su robustez en la transformación de datos y la generación de informes dinámicos.

Se diseñaron funciones reutilizables para cada etapa del pipeline, se estandarizaron las estructuras de datos requeridas por tipo de reporte, y se incorporaron mecanismos de monitoreo en tiempo real. Esto incluyó la generación de logs automáticos, alertas de error o finalización enviadas por correo electrónico y Telegram, y validaciones en puntos críticos del flujo.

El sistema actualmente permite ejecutar reportes en los 34 hospitales generales y especializados, y en los 51 Centros de Salud y Acción Comunitaria (CeSACs), adaptando la lógica a las particularidades institucionales. La implementación de este enfoque permitió asegurar la consistencia de los resultados, facilitar su mantenimiento, y dotar al proceso de trazabilidad, escalabilidad y capacidad de auditoría.

4 Resultados

El nuevo pipeline fue implementado entre noviembre de 2024 y abril de 2025, e incluyó la automatización total de los reportes mensuales en la totalidad de los efectores de la Red. Los resultados fueron alentadores: el tiempo promedio de generación de documentos se redujo de 30 a 9 horas, lo que representa una disminución superior al 60%. Además, el nuevo enfoque optimiza el uso de memoria al procesar los datos de forma secuencial por día, en lugar de ejecutar todo el procesamiento en una única corrida diaria. Esta mejora permitió eliminar la necesidad de intervención manual, al evitar fallas recurrentes en la ejecución.

Se mejoró también la disponibilización de los reportes para el área de facturación, lo que permitió acelerar la gestión ante los financiadores. Si bien no se presentan en este artículo métricas de impacto económico específicas, se estima que la reducción en tiempos operativos y la mejora en la oportunidad de presentación de documentación contribuyen de manera significativa a fortalecer la capacidad de recupero de gastos del subsistema público de salud de CABA¹. Adicionalmente, se consolidaron

¹ Los datos de impacto económico y facturación vinculados al recupero de gastos son de carácter interno del Ministerio de Salud de la Ciudad Autónoma de Buenos Aires. Por este motivo, en el presente artículo se presentan únicamente métricas relacionadas con la cantidad de prestaciones procesadas y los tiempos operativos, resguardando la confidencialidad institucional.

indicadores de trazabilidad interna que permiten monitorear en tiempo real el estado de cada proceso y detectar puntos de falla. Esta información resultó clave para la toma de decisiones operativas, facilitando la priorización de tareas, la asignación de recursos y la identificación de necesidades de formación en los equipos locales.

5 Consideraciones finales

El proyecto se posiciona como una solución escalable y replicable, con alto potencial para ser adoptada por otras jurisdicciones. Las líneas de trabajo futuras apuntan a profundizar su valor agregado mediante la incorporación de nuevas fuentes de datos, el desarrollo de herramientas visuales de apoyo a la gestión, y la evaluación sistemática de su impacto desde una perspectiva técnica, operativa y de política pública.

En este sentido, la experiencia adquirida durante la fase piloto sienta las bases para una transformación más amplia del uso de datos en el ámbito de la salud pública, promoviendo procesos más ágiles, transparentes y basados en evidencia.

Referencias

- Averin, P., Mellidou, I., Ganopoulou, M., Xanthopoulou, A., & Moysiadis, T. (2025). Evaluating Directed Acyclic Graphs with DAGMetrics: Insights from Tuber and Soil Microbiome Data. *Agronomy*, 15(4), 987. <https://doi.org/10.3390/agronomy15040987>
- Castejón-Limas, M., Fernández-Robles, L., Alaiz-Moretón, H., Cifuentes-Rodriguez, J., & Fernández-Llamas, C. (2022). A Framework for the Optimization of Complex Cyber-Physical Systems via Directed Acyclic Graph. *Sensors*, 22(4), 1490. <https://doi.org/10.3390/s22041490>
- Dalianis, H. (2018). *Computational methods for text analysis and text classification*. H. Dalianis (Ed.), *Clinical Text Mining: Secondary Use of Electronic Patient Records* (pp. 83–96). Springer. https://doi.org/10.1007/978-3-319-78503-5_6
- Gobierno de la Ciudad Autónoma de Buenos Aires. (s.f.). Facturación y cobranza de efectores públicos - Institucional. Recuperado de <https://buenosaires.gob.ar/salud/facturacion-y-cobranza-de-efectores-publicos/institucional> (último acceso: 2025/04/14).
- Gobierno de la Ciudad Autónoma de Buenos Aires. (2022). Resolución N.º 2726/MJGGC/22. Boletín Oficial de la Ciudad de Buenos Aires. Recuperado de <https://boletinoficial.buenosaires.gob.ar/normativaba/norma/332530> (último acceso: 2025/04/14).
- Nanton, M. C., Bardauil, A., & Rodríguez Tablado, M. (2024). *Experiencias en el uso de R para la automatización de procedimientos administrativos: el recupero de gastos en el sistema público de salud de la Ciudad de Buenos Aires*. LatinR 2024: Conferencia Latinoamericana sobre Uso de R en Investigación + Desarrollo.
- Shubha, B. G., & Prasad, A. M. (2019). Airflow Directed Acyclic Graph. *Journal of Signal Processing*, 5(2), 12–20. <http://doi.org/10.5281/zenodo.3247274>
- Tennant, P. W. G., Murray, E. J., Arnold, K. F., et al. (2021). Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International Journal of Epidemiology*, 50(2), 620–632. <https://doi.org/10.1093/ije/dyaa213>