



Comisión
Nacional de
Productividad

CALIDAD DE DATOS Y SISTEMAS DE INFORMACIÓN EN SALUD PÚBLICA

NOTA TÉCNICA
ENERO 2020

Autores:

Valeria Lobos Ossandón. *Data Scientist, Comisión Nacional de Productividad.*

Andrés Olivares Arredondo. *Computer Scientist, Comisión Nacional de Productividad.*

Agradecemos a las siguientes personas por su valioso aporte en la confección de este informe:

Mónica Cortés. *Asesora de diseño de servicios complejos.*

Érika Quiroz. *Asesora de diseño de servicios complejos.*

Alejandro Mauro. *Jefe del Departamento de Informática Biomédica, Clínica Alemana, Santiago.*

Tabla de contenido

Introducción.....	4
Ficha Clínica Electrónica: El Eje Central de Datos	5
Experiencias Internacionales del Uso Ejemplar de los Datos y Tecnología.....	7
UK National Health Systems	7
Johns Hopkins Capacity Command Center.....	8
Estudio del Corazón Framingham	9
Datos y Sistemas de Información en el Sistema de Salud Público Chileno	10
Estadísticas Sanitarias en Chile	10
Departamento de Estadísticas e Información de Salud.....	10
Sistema de Información de Salud Maule.....	12
Proyecto SIDRA	12
Hospital Digital	13
Registro Único de Lista de Espera	15
Proyecto Almohadita – Ámbar	15
Arquitectura de Software en MINSAL	16
Análisis a las Bases de Datos del Sistema de Salud Público	20
Aspectos Claves de las Bases de Datos	20
Descripción de las Bases de Datos	21
Indicadores de Calidad de Datos	22
Accesibilidad	23
Cantidad de Datos Apropriados	25
Concisión.....	26
Consistencia	28
Compleitud	29
Credibilidad	31
Facilidad de Manipulación	32
Interpretabilidad	33
Libre de error.....	34
Temporalidad.....	37
Seguridad.....	39
Conclusión.....	42
Trabajo Futuro.....	43
Bibliografía	44

El presente estudio se emite en el marco del proyecto de Eficiencia en Pabellones, encomendado por la Presidencia a la Comisión Nacional de Productividad. Este documento tiene por objetivo realizar una evaluación de la calidad de datos y sistemas de información del sistema de salud en Chile.

Introducción

Los datos son indispensables para la realización de análisis, y se transforman en información útil para la toma de decisiones y en conocimiento. Por lo tanto, si los datos no son correctos, las conclusiones que surjan a partir de ellos serán equívocas también. Esto resalta la imperiosa necesidad de contar con datos fidedignos, trabajarlos correctamente y realizarles una limpieza adecuada, para trabajar con ellos y llegar a conclusiones correctas.

Actualmente, los datos en salud pública son suficientes para tomar decisiones, pero requieren de un costoso procesamiento y un costo en precisión y calidad. Por lo tanto, existen oportunidades de mejora que permitirían llegar a conclusiones más precisas. Este documento busca indagar sobre la validez y certeza de los datos en salud pública, con el fin de mejorar la calidad de datos e información del sistema de salud chileno.

Junto con lo anterior, se destaca también el aporte de diversos estudios que han señalado la baja calidad de los datos y sistemas de información en salud, los cuales necesitan una revisión (Centro de Políticas Públicas UC, 2017; Contraloría, 2018; Contraloría, 2019). Una de las motivaciones de este documento es complementar y cuantificar dichos estudios.

La inclusión de sistemas de información y tecnología trae ventajas para todos los involucrados en el proceso de atención de los pacientes. Éstos se ven beneficiados por una mejor atención, disminución de la burocracia y un mayor empoderamiento del conocimiento de su estado de salud. El personal médico se beneficia de la tecnología al tener una rápida disponibilidad del estado de salud del paciente, lo que le permite tomar mejores decisiones. También, la tecnología brinda al personal médico de una rica fuente de datos para realizar investigaciones. Por último, los sistemas de información son un apoyo de gestión para el personal administrativo.

Esta nota técnica tiene por objetivo hacer una evaluación a los datos del sistema de salud público chileno y a sus sistemas de información, y también proponer soluciones a futuro. El documento analiza cada una de las bases de datos a la cual tuvo acceso la Comisión y se presenta un análisis cuantitativo con indicadores de calidad de datos, para determinar qué tan buenos son los datos en salud.

Ficha Clínica Electrónica: El Eje Central de Datos

La Ficha Clínica Electrónica es el eje central en cuanto a datos de salud se refiere. Es un instrumento que registra el historial del estado de salud del paciente en formato digital, lo cual permite obtener una visión acabada de su estado (Office of the National Coordinator for Health Information Technology, 2019). El objetivo principal de la Ficha Clínica Electrónica es dar soporte a la toma de decisiones médicas.

La Ficha Clínica Electrónica comprende, en general, los siguientes aspectos del paciente: datos de contacto, información acerca de sus citas médicas, información con respecto a su plan de salud, historia familiar, alergias, historial de vacunas, historial de diagnósticos y comorbilidades, historial de medicamentos, hospitalizaciones y cirugías o procedimientos realizados, entre otros (TechTarget, 2017). De este modo, engloba toda información que tienen directa relación con la salud histórica del paciente.

Adicionalmente, la Ficha Clínica Electrónica también comprende un amplio espectro de otros componentes, tales como aspectos legales y éticos, estándares de notación, interoperabilidad de registros y los metadatos (Kalra, 2016). Además, estudios mencionan que sólo el 40% de los datos en salud está efectivamente en la Ficha Clínica Electrónica, mientras que el 60% restante se encuentra en distintas bases de datos, como en las farmacias, datos genéticos, exámenes solicitados, entre otros (Weber, Mandl, & Kohane, 2014; Diana, 2014; Mauro, 2019).

La Ficha Clínica Electrónica tiene múltiples ventajas con respecto al mismo instrumento en formato físico (en papel). Dentro de estos beneficios se incluye la disminución en tiempos de documentación, incremento en la seguridad con la cual se almacena y se accede a ella, facilidad de acceso desde distintas locaciones, incremento de la facilidad de interpretación de la ficha, entre otros (Kalra, 2016). La Ficha Clínica Electrónica también puede ayudar a que se comparta de forma fácil y segura el historial médico de los pacientes entre servicios de salud, así como también puede incluir atributos calculados con analítica, que aporten a una mejor comprensión del estado de salud del paciente (por ejemplo, se puede incluir el riesgo sanitario del paciente). En efecto, la Ficha Clínica Electrónica definitivamente contribuye a mejorar la calidad de atención a los pacientes (Hillestad, y otros, 2005; Fisher, Bhavnani, & Winfield, 2009; King, Patel, Jamoom, & Furukawa, 2014).

La Ficha Clínica recién fue regulada en Chile en el año 2012, pero su regulación no incluye la obligatoriedad de que sea en formato electrónico. Esto implica de que las instituciones de salud no están obligados a implementar la Ficha Clínica Electrónica.

La implementación de la ficha, así como la implementación de cualquier software, requiere de una fuerte inversión monetaria en tecnología. En el año 2009, se hizo un desembolso monetario de 19 billones de dólares en USA para implementar y fomentar el uso de la Ficha Clínica Electrónica (Blumenthal, 2009). Sin embargo, la implementación de la ficha trae consigo la disminución de costos administrativos a largo plazo (Wang, y otros, 2003). De acuerdo a datos de DIPRES, en Chile se desembolsa cerca de un 1% del presupuesto de salud en tecnología.

Desde la perspectiva legal la Ficha Clínica Electrónica no está regulada (FALMED, 2019). Sin embargo, la Ficha Clínica Electrónica es una versión digital de la Ficha Clínica, la cual sí está regulada. Este es un instrumento obligatorio que puede estar en formato digital o en papel, que registra el conjunto de antecedentes relativos a las diferentes áreas relacionadas con la salud de una persona, cuya finalidad es mantener integrada la información necesaria para el otorgamiento de atenciones de salud (Varas, Hering, & Demetrio, 2010; Decreto N°41, 2012). Adicionalmente se agrega que en este documento debe quedar registrada toda actuación profesional que el equipo de salud brindó a un paciente, como por ejemplo procedimientos realizados o los nombres de profesionales que atendieron al paciente.

Asimismo, desde la perspectiva legal se ha señalado que constituye la historia clínica del paciente y el reflejo del método clínico, además de un instrumento guía para todo el quehacer sanitario, estadístico e investigativo (Godoy & Barraza, 2018) que pueda suscitarse en el contexto de las prestaciones asistenciales. Este documento de carácter obligatorio debe ser redactado por el profesional de la salud que realiza la evaluación, efectuándose un seguimiento de las consultas y tratamientos proporcionados. La historia clínica es uno de los componentes fundamentales para habilitar el modelo operacional de la estrategia de Redes Integradas al Servicio de Salud (RISS) y la continuidad del cuidado.

En Chile, la Ficha Clínica Electrónica es parte del programa de implementación SIDRA, del cual se detalla más adelante en este documento.

Experiencias Internacionales del Uso Ejemplar de los Datos y Tecnología

La Ficha Clínica Electrónica es la fuente principal de datos de salud e implementada correctamente, ayuda a mejorar la calidad de la atención de los pacientes y es un soporte a las decisiones clínicas. Por ejemplo, en el caso del Hospital Johns Hopkins, una implementación correcta de la Ficha Clínica mejoró en un 60% las capacidades para admitir pacientes complejos, mientras que el estudio del corazón Framingham muestra la importancia del historial clínico como herramienta para estimar el riesgo de los pacientes. A continuación, se presentan experiencias internacionales en el uso de la Ficha Clínica Electrónica.

UK National Health Systems

NHS (National Health Systems) es el sistema de tecnología e información del sistema nacional británico para mejorar su sistema de salud. Su misión es utilizar la información y tecnología para mejorar el sistema de salud (NHS, 2014; NHS, 2019). La información recopilada por el sistema es luego utilizada en publicaciones científicas que aportan valor al sistema de salud. Producto de este trabajo han surgido innovadores indicadores en distintas áreas, como enfermedades infecciosas, salud infantil, mortalidad, embarazo, entre otras (NHS, 2018; NHS, 2019; Topol, 2019)

El NHS se responsabiliza de recopilar y estandarizar datos de todo el sistema de salud británico. Facilita la interoperabilidad en toda la red, independiente de dónde se haya atendido el paciente. La institución utiliza el principio de que malos datos llevan a malas decisiones, por lo cual ha implementado una serie de estándares y buenas prácticas, y ha entrenado al personal médico sobre cómo almacenar los datos correctamente.

Pese a lo anterior, el NHS entiende que el tiempo del personal médico es limitado como para utilizarse en tipear datos, por lo cual gran parte de los datos recolectados se canalizan por medio de la Ficha Clínica Electrónica y el Open Exeter (NHS, 2018). Esta última es una herramienta que facilita el ingreso de datos y brinda acceso a la información del paciente, como exámenes ginecológicos, información de donante de órganos, tipo de sangre, entre otros.

Una vez recolectados los datos, el sistema se encarga de compilarlos. En esta etapa se da feedback a las diversas instituciones con respecto a la calidad de los datos, y además se asegura de que se cumplan los estándares de privacidad de los datos (NHS, 2019).

En las etapas finales del manejo de datos, el NHS realizan análisis a partir de los datos y calcula distintas estadísticas, las cuales distribuye posteriormente al sistema, para que se tomen mejores decisiones.

Por último, se destaca también que el NHS ha implementado un sistema nacional de agendamiento ambulatorio, de exámenes y de hospitalizaciones. Además, dispone de un portal para pacientes que promueve activamente que los pacientes se involucren en la atención de salud

Johns Hopkins Capacity Command Center

En USA, el Hospital Johns Hopkins ha implementado una Torre de Control (Capacity Command Center), cuyo objetivo es entender el flujo del paciente por el hospital. La Torre de Control asimila sus funciones a una torre dentro de un aeropuerto, pero en vez de controlar vuelos, controla el flujo de los pacientes por el hospital.

Básicamente, la Torre de Control recopila datos en tiempo real para tomar mejores decisiones. Utiliza principios de ingeniería, pero aplicados a la medicina. Su infraestructura tecnológica funciona como un repositorio de datos que luego son usados para generar modelos predictivos y proponer soluciones creativas para diversos problemas (Johns Hopkins, 2016). La Torre integra diferentes fuentes de bases de datos, como por ejemplo resultados de exámenes del pacientes, diagnósticos, así como también datos de horarios de personal médico (Roth, 2018).

En la Torre de Control trabajan en conjunto miembros de diferentes departamentos monitoreando datos en tiempo real y análisis predictivos. La Torre permite identificar cuellos de botella, reducir tiempos de espera, coordinar servicios y reducir riesgo. Asimismo, también implementa alertas en tiempo real que requieren tomar acciones inmediatas (Johns Hopkins, 2016).

La tecnología de la Torre de Control mantiene al personal informado durante todo el tiempo con respecto al estado en tiempo real del hospital. Por ejemplo, informa dónde está un paciente específico en el hospital, qué unidades necesitan personal de apoyo, el número de pacientes que están siendo atendidos, la disponibilidad de camas en el hospital, el número de ingresos y altas de pacientes, entre otros.

La evaluación de la implementación de la Torre de Control ha sido positiva. Dentro de sus principales logros, se incluyen:

- Mejora en un 60% de la habilidad para admitir pacientes complejos.
- Disminución en 63 minutos del tiempo que demora una ambulancia en trasladar pacientes entre hospitales.
- Los pacientes que ingresan a urgencias y son hospitalizados son asignados a una cama un 30% más rápido.
- En un 70% se han reducido los tiempos de espera entre pabellones.
- Se ha incrementado en un 21% el número de pacientes que son dados de alta antes del mediodía.

La Torre de Control no sólo ayuda a mejorar procesos clínicos, sino que también es un apoyo a la atención sanitaria. Los datos recopilados son utilizados por un grupo multidisciplinario de científicos para crear modelos predictivos, como por ejemplo predicción de riesgo de cierta enfermedad o categorizar el riesgo sanitario de un paciente, entre otros.

Estudio del Corazón Framingham

El Framingham Heart Study es un proyecto que busca identificar los factores de riesgo que incrementan el riesgo de contraer una enfermedad cardiovascular. El estudio comenzó en el año 1948 con una cohorte de 5.209 personas, a las cuales ha seguido a lo largo de sus vidas para identificar factores de riesgo. A lo largo de los años, se han incorporado a los descendientes de los participantes originales del grupo (Framingham Heart Study, 2019).

Los datos anonimizados de los pacientes pueden ser solicitados por investigadores, con el objetivo de generar conocimientos al respecto. El uso de información de este tipo ha facilitado nuevos descubrimientos con respecto a qué variables deben revisar los médicos para prevenir un episodio cardiovascular, así como la creación de indicadores más precisos para anticiparse a enfermedades cardiovasculares.

El estudio Framingham es un ejemplo de cómo el trabajo sobre una base de datos sanitaria robusta trae consigo importantes avances en cuanto a conocimiento médico.

Datos y Sistemas de Información en el Sistema de Salud Público Chileno

La siguiente sección tiene por objetivo hacer un catastro de los principales proyectos y aplicaciones de tecnología en el sistema de salud público. A continuación, se describe el contexto nacional de los sistemas de información y datos en salud pública. Se resaltan ejemplos nacionales públicos del buen uso de tecnologías y se mencionan iniciativas ministeriales en dicha línea.

Estadísticas Sanitarias en Chile

El Ministerio de Salud, en el marco legal de sus funciones, es el encargado de tratar datos con fines estadísticos y mantener registros y bancos de datos respecto de las materias de su competencia. Para esto, se encuentra en facultad para tratar datos personales o sensibles con el fin de proteger la salud de la población o para la determinación y otorgamiento de los beneficios de salud (Decreto con Fuerza de Ley N°1, 2005).

A mayor abundamiento, el reglamento orgánico del Ministerio de Salud precisa, que este tiene el deber de diseñar, implementar y mantener actualizados sistemas de información que permitan proporcionar datos estadísticos (Decreto N°136, 2005) para:

- La formulación, control y evaluación de programas de salud.
- El desarrollo de infraestructura en salud.
- La gestión de recursos humanos y financieros.
- La producción del sector.
- Los impactos directos que sus acciones en salud generen en el estado de salud de la población.
- La calidad de atención de los pacientes.

Asimismo, se señala que estos sistemas de información abarcarán materias tales como datos estadísticos de morbilidad, mortalidad y sus causas específicas, datos acerca de la oferta de prestaciones de salud, coberturas de atención, salud ambiental, ocupacional y todas aquellas que fueren necesarias para el cumplimiento de sus cometidos legales.

El Ministerio de Salud, creó mediante la resolución exenta N°2.164 del año 2000 el Departamento de Estadísticas e Información de Salud (DEIS), al cual este informe pasa a referirse.

Departamento de Estadísticas e Información de Salud

Este departamento está subordinado a la División de Planificación Sanitaria (DIPLAS) que a su vez forma parte de la Subsecretaría de Salud Pública (Decreto N°136, 2005).

En lo concerniente a esta materia, la referida subsecretaría tiene la función de proponer las políticas de información y de tecnologías de la información del sector y velar por su aplicación. Además, está encargada de proponer los estándares de datos, de códigos y clasificaciones en los sistemas de información del sector, guardando por su observancia.

En cuanto a la información pública que provee DEIS en su sitio web oficial, se dispone de las siguientes publicaciones:

- Indicadores básicos de salud para los años 2009 a 2016.
- Estadísticas de natalidad y mortalidad para los años 2007 a 2014.
- Estadísticas de Atenciones y Recursos para la Salud para los años años 1999 a 2009.

Sin perjuicio de lo anterior, la DIPLAS es la entidad encargada de ejercer la jefatura sobre DEIS, toda vez que las funciones de este departamento se engloban en uno de los cometidos institucionales de DIPLAS. Es decir, es la encargada de producir sistemática y oportunamente planes, normas, estadísticas e información en respuesta a la demanda de prestaciones y en forma coherente con las políticas del Sector (MINSAL, 2018), en ejercicio de las facultades de la Subsecretaría de Salud Pública.

Así las cosas, DEIS, en tanto departamento de DIPLAS, es el encargado de hacer efectiva la obligación del Ministerio de Salud de tratar los datos sanitarios de la población con fines estadísticos, en aras a que el Ministerio pueda formular, fijar y controlar las políticas de salud, y ejercer la rectoría sectorial. Es por esto que DEIS cuenta con las siguientes funciones (Unidad de Estadísticas y Estándares, DEIS, 2011):

- Producir la información estadística oficial del Ministerio de Salud.
- Difundir la información del Sector en forma de publicaciones.
- Mantener información oficial acerca de temas de salud así como de sus factores relacionados de modo que se encuentre disponible para los distintos tipos de usuarios.
- Administrar la información proveniente de los organismos del Sector Salud, vigilando la integridad, oportunidad, exactitud y comparabilidad de los datos.

Similar al NHS, se sugiere que el DEIS elabore estadísticas e índices de rendimiento a nivel nacional, y que luego envíe dichas estadísticas a los centros hospitalarios, con el fin de que cada establecimiento tenga un estándar para compararse. Se propone que la información enviada a los establecimientos esté presentada de forma didáctica y fácil de entender, para su fácil interpretación.

- Proporcionar la información necesaria para la gestión del sector, apoyando la toma de decisiones.
- Establecer los estándares de información estadística y ponerlos a disposición del sector Salud.
- Dictar normas y procedimientos relacionados con la recolección, procesamiento, presentación y distribución de información.
- Mantener series históricas de estadísticas del sector salud.
- Promover y coordinar actividades de capacitación del personal que desempeña funciones relacionadas con la producción de información y estadísticas de salud.

Sistema de Información de Salud Maule

El SISMAULE (Sistema de Información de Salud Maule) es un sistema de información de la región del Maule que comprende a hospitales, CESFAM, consultorios y postas. El sistema apunta a la creación de esta Ficha Clínica Electrónica a nivel regional (SSMAULE, 2014).

SISMAULE da soporte a la región con respecto al registro clínico, estadísticas de atención primaria de salud, estadísticas hospitalarias, categorización de pacientes hospitalizados, datos de la población bajo control, además de la lista de espera (Salud-e, 2015).

La principal fortaleza de este sistema de información, es que mantiene centralizada una misma base de datos a nivel regional. Esto permite un registro único regional de Ficha Clínica Electrónica y de profesionales. Si bien SISMAULE es un buen ejemplo de la implementación de la Ficha Clínica nacional, se debe destacar que es difícil escalar un sistema de este tipo a todo el país, dada la variabilidad en tecnologías que se utilizan en los distintos establecimientos de salud.

Proyecto SIDRA

SIDRA (Sistema de Información de Redes Asistenciales) es un proyecto de tecnologías de información que automatiza la información clínica (SSMC, 2019). Provee equipamiento y sistemas de información a toda la red hospitalaria (Revista Gerencia, 2012). El proyecto parte de la agenda de digitalización del Gobierno, y además busca que para el año 2020 todos los chilenos tengan una Ficha Clínica Electrónica (Salud-e, 2016), bajo estándares internacionales HL7 (Health Level Seven) (HL7, 2019).

En específico, SIDRA es responsable de los siguientes módulos:

- **Sistema de lista de espera:** encargado de derivar pacientes en la red de salud. Si bien este era uno de los objetivos del sistema de lista de espera, hoy en día no es más que un sistema de archivos, es decir, sólo almacena datos y no se encarga de su gestión.
- **Agenda:** sirve para programar recursos, incluyendo personas, equipo y suministros.
- **Urgencias:** automatiza y estandariza el registro de pacientes en entornos de cuidados de urgencias.

- **Farmacia:** gestiona el inventario de medicamentos de cada hospital.
- **Programa registro de población bajo control:** aprueba y promueve programas de salud implementados por el gobierno.

En la práctica, el proyecto SIDRA carece de estrategia: no abordó procesos sino que introdujo tecnología a tareas independientes. De este modo, solamente digitalizó ciertos procesos, generando soluciones silos que no son parte de una estrategia de informatización del Ministerio (Cortés & Quiroz, 2019).

SIDRA comenzó en el año 2008 con la gestión de listas de espera, y en el año 2012, año en que se regula la Ficha Clínica Electrónica en Chile, tuvo a cargo la implementación de la Ficha. En el año 2016, existieron cambios en cuanto a la licitación del proyecto SIDRA II, el cual se suspendió temporalmente porque no incluía ciertos cambios en la normativa de ciberseguridad (Economía y Negocios, 2017). Sin embargo, el plan de implementar SIDRA y digitalizar e interconectar a los distintos sistemas de información en salud seguía presente en el año 2018 (Salud-e, 2018). Asimismo, al presente año 2019, se está trabajando en implementar la infraestructura que dará soporte a la Ficha Clínica Electrónica en Chile y a la interoperabilidad de esta (CENS, 2018), pero no como parte del proyecto SIDRA. En la actualidad, se desconoce el avance del proyecto SIDRA como tal.

Hospital Digital

El Hospital Digital es un modelo de atención en salud, a cargo del Ministerio de Salud, que tiene por objeto poner en marcha y operar soluciones de medicina digital a distancia, para proveer un acceso oportuno y de calidad en atenciones de salud, aumentando las coberturas horarias de atención, contribuyendo a la reducción de listas de espera y mejorando los niveles de resolutivez de la atención primaria en salud (Resolución Exenta N° 347, 2019). Según fuentes del MINSAL, Hospital Digital actualmente se centra en determinar la oferta médica, pero no así la demanda.

Esta iniciativa fue concretada, por medio de una plataforma en internet, en mayo del 2019 (Hospital Digital, 2019), y se enmarca dentro de la implementación de la estrategia digital del sector salud. Esta estrategia tiene por misión contribuir a mejorar la salud de la población a través de gestión oportuna y eficiente de información estandarizada. De este modo, la estrategia digital está integrada por una serie de propuestas que incorporan el uso de tecnologías de la

información y comunicaciones como herramientas esenciales para la modernización de la gestión en salud (Salud-e, 2019).

En esta línea, la Subsecretaría de Redes Asistenciales recientemente estableció, el Departamento de Salud Digital (Resolución Exenta N° 595, 2019), el cual mediante el uso de tecnologías de la información y comunicación, propende a facilitar la provisión de servicios de salud a distancia. Es el organismo encargado de implementar el proyecto de Hospital Digital y coordinar las acciones destinadas a su desarrollo. Dentro de los organismos que lo componen está el Departamento de Hospital Digital, estrechamente vinculado con el Departamento de Tecnología de la Información.

El Departamento de Tecnología de la Información es parte de MINSAL y sus funciones son las siguientes:

- Gestionar los recursos, infraestructura y servicios tecnológicos institucionales mediante la mantención y desarrollo de sistemas de información y servicios informáticos.
- Velar por el cumplimiento de los objetivos de la estrategia de tecnologías de la información en salud.
- Dirigir los proyectos tecnológicos del sector con énfasis en la integración de la información entre los diferentes niveles asistenciales para la optimización de la gestión y registros oportunos y de calidad.
- Brindar asesoría al Ministerio y a sus subsecretarías en las materias de tecnología de la información y comunicaciones.
- Garantizar la interoperabilidad e intercambio de información.

En cuanto al Departamento de Hospital Digital, este fue el encargado de diseñar, construir, poner en marcha y operar las soluciones de medicina digital a distancia. Asimismo, este departamento puede dividirse en tres líneas de trabajo.

- **Telemedicina:** Esta unidad se ocupa de los procesos de, atención ambulatoria, atención de hospitalización, atención de urgencia, procesos quirúrgicos y de apoyo de diagnósticos y terapéuticos.
- **Teleeducación:** Esta unidad se centra en los aspectos de teleinformación, telecapacitación y telepromoción.
- **Teleasistencia:** Esta unidad se encarga del monitoreo, seguimiento y orientación a distancia.

Adicionalmente, el Departamento de Hospital se subdivide orgánicamente en las siguientes oficinas:

- Oficina de Red Hospital Digital
- Oficina de Monitoreo de Prestaciones
- Oficina de Monitoreo Tecnológico

- Oficina de Telemedicina.

Desde la perspectiva jurídica, existe una duplicidad de organismos, dado que existe un departamento de Hospital Digital dependiente del Departamento de Salud Digital, y otro departamento del mismo nombre dependiente del Departamento de Tecnología de la Información, dependiente del Ministerio de Salud.

Lo señalado se observa al contrastar la información de dos resoluciones exentas del año en curso (2019): la resolución exenta n°347 del 21 de febrero elaborada por la Dirección Jurídica del Ministerio de Salud, y resolución exenta n°595 del día 02 noviembre, emanada de la Subsecretaría de Redes.

Sin perjuicio de lo anterior, el Departamento de Hospital Digital dependiente del Departamento de Tecnología de la Información tiene (o tuvo) por cometido el diseño, construcción, puesta en marcha y operación de las prestaciones asociadas a la medicina a distancia. En cambio, el Departamento de Hospital Digital que forma parte del Departamento de Salud Digital, es el encargado de facilitar la provisión de servicios de salud a distancia. Bajo estos antecedentes, operar y facilitar la provisión de prestaciones a distancia, constituye una duplicidad de funciones.

Por último, se destaca que la iniciativa está siendo investigada por Contraloría por inconsistencias en algunos gastos (Cluster Salud, 2019).

Registro Único de Lista de Espera

En la cuenta pública de 2018 se anunció la creación del Registro Unificado de Pacientes que mantendrá en un único repositorio todos los pacientes en lista de espera (Mauro, 2018). Concretamente, esto corresponde a dos sitios webs. El primero de ellos presenta la lista de espera quirúrgica NO GES actualizada al día (MINSAL, 2019), mientras que el segundo muestra un monitoreo ciudadano de listas de espera quirúrgica no GES (MINSAL, 2019).

Ambos sitios tienen por objetivo transparentar las atenciones de lista de espera. Sin embargo, a octubre del año 2019 se ha constatado que la lista de espera quirúrgica NO GES no se encuentra disponible en el sitio web.

Proyecto Almohadita – Ámbar

Los proyectos Almohadita y Ámbar (Almohadita, 2019) son proyectos de colaboración multidisciplinaria en los cuales se utilizan datos médicos personales en conjunto con técnicas estadísticas y de *machine learning*, para observar de forma fácil el estado de salud de pacientes infantiles.

El proyecto Almohadita es un sistema de monitoreo a distancia capaz de anticipar crisis respiratorias, el cual fue implementado en el Hospital Dr. Exequiel González Cortés (Ríos, 2019). El proyecto nació en el año 2015 financiado por un concurso FONDEF, en colaboración del hospital con el Departamento de Ingeniería Industrial (DII) de la Universidad de Chile. El sistema captura los signos vitales de los pacientes de dos maneras: a) por medio de una ficha electrónica diseñada para tablets, celulares y PCs. b) Por medio de una plataforma de sensores que permite hacer la captura de datos de manera automática.

El proyecto Ámbar consiste en un software que monitorea en tiempo real el estado de salud de lactantes, y utiliza los datos para predecir su evolución del estado de salud. De esta forma, alerta de posibles complicaciones que el paciente pudiese sufrir (El Mercurio, 2016).

Ambos sistemas son de uso gratuito en el hospital, y está disponible gratuitamente para cualquier hospital que requiera implementarlo (Ríos, 2019).

En la actualidad, el Proyecto Ámbar creció para convertirse en la Ficha Clínica Electrónica de los pacientes del Hospital Exequiel González Cortés, la cual permite el registro y acceso a las consultas, evaluaciones, interconsultas, derivaciones y requerimientos en las atenciones de urgencia, de hospitalizados, quirúrgica y ambulatoria.

El proyecto Almohadita – Ámbar es un ejemplar caso nacional de éxito resultante de la colaboración multidisciplinaria, en pos de ayudar a mejorar la calidad de atención de los pacientes.

Arquitectura de Software en MINSAL

La arquitectura de software corresponde a la estructura fundamental sobre la cual se crea un sistema de información. La elección de arquitectura debe realizarse una vez definido el problema en su totalidad e identificado lo que se busca resolver. Es fundamental tener una amplia visión del uso que se le dará al software y de las posibles nuevas funciones que debería cumplir la herramienta en un futuro.

La decisión de la arquitectura de software involucra tomar en cuenta sistemas de información externos e internos, tanto herramientas y datos que existían previamente como otras herramientas creadas por entidades con las que el sistema se comunicará o interactuará eventualmente. Esta decisión también debe considerar que los sistemas de información en salud deben aportar a la salud del paciente. Por ejemplo, existe una dualidad entre el usuario y el cliente, donde el cliente espera una cosa del sistema, pero los usuarios esperan otra. En la práctica el MINSAL es un cliente y los médicos son usuarios de un sistema. Si los médicos no utilizan el sistema (porque no les ayuda o porque no hay incentivo en su uso) este no cumplirá su función. Por otro lado, si el sistema solo realizara lo que ellos quieren, este podría no ser de utilidad para el MINSAL, que necesita más datos para tomar decisiones.

Una arquitectura orientada a servicios es una arquitectura en la que un sistema se compone de múltiples piezas que responden a consultas originadas por otra componente de forma estandarizada. En la práctica, esta permite poner la información a disposición de distintas entidades y/o aplicaciones sin generar cambios en sus funciones.

Para enfrentar los principales problemas de propiedad, tanto intelectual como de datos, y las dependencias que se derivan de las licitaciones, es que se sugiere utilizar tecnologías *open source* o código abierto. Lo anterior facilita las tareas de cambiar de empresa, realizar cambios sobre el software y manejar los datos sin depender del proveedor original.

MINSAL tuvo en sus planes implementar un sistema de información para una estrategia nacional de salud, basada en arquitectura de software orientada a servicios (Equipo IBM, 2016). Sin embargo, según fuentes internas del Ministerio, el sistema de datos que cumplía con esta característica no sólo está en desuso, sino que también está obsoleto. Esto es un ejemplo similar al proyecto de software auditado por Contraloría (2019), en el cual se desembolsó US\$4,8 millones y que finalmente no se utilizó. Ambas experiencias, dan cuenta de

que se hacen grandes inversiones en proyectos de software que finalmente son inútiles.

Sumado a lo anterior, una particularidad de la arquitectura orientada a servicios es que requiere un esfuerzo conjunto de las unidades independientes que disponen de ellos, traducido en la constante atención y mejora de los propios sistemas o componentes para responder a las actualizaciones y nuevas funcionalidades que elaboran otros. Tal compromiso añade complejidad humana al funcionamiento del sistema, que es una de las grandes falencias de este tipo de arquitectura.

La complejidad de la arquitectura de software orientada a servicios, también puede verse afectada cuando una gran cantidad de los componentes dependen de externos contratados, pues se debe solicitar un requerimiento que no estaba planteado en el software inicialmente. Dado que solo el dueño del software (la empresa externa) puede realizar este cambio, el precio es definido por ellos mismos. Un ejemplo de esto es el sistema de información de inmunización contratado por el MINSAL, para el cual el contrato no le permitía al Ministerio acceder a los datos, y cada información adicional a la estipulada por el contrato, era cobrada de forma adicional (BID, 2019).

La arquitectura va de la mano con la metodología de desarrollo, que corresponde a la estrategia de cómo se producen las piezas de software, añadiendo una dimensión adicional en la dificultad al decidir el tipo de cimientos que tendrá un sistema que funcionará durante un largo período de tiempo.

Dado el constante cambio y la polarización política histórica entre los servicios públicos del área, no existe una comunicación efectiva entre dichos servicios. Un cambio interno de sistema o de personal puede cambiar tanto acuerdos realizados como protocolos establecidos que se utilizan para crear una solución. Sin embargo, existen metodologías que permiten lidiar con problemas de este tipo con mayor facilidad que otras. Este tipo de “Metodologías ágiles” permiten mejorar constantemente versiones entregables del producto donde agregar o quitar funcionalidades no implica un costo tan alto (como lo sería con metodologías de grandes sistemas definidos como la metodologías de cascada).

Por último, se destaca también los resultados de documento de auditoría a MINSAL, en el cual se señala que existe un desconocimiento tanto del software como hardware que existen en el Ministerio (Contraloría, 2019). Por lo anterior, se sugiere realizar un catastro del software y hardware. Además, se propone hacer una revisión de pagos por licencias de software que no se están usando. También se sugiere imitar las buenas prácticas de países europeos como Francia y dejar de comprar licencias de software cuyo uso habitual sólo contempla funciones que pueden ser reemplazadas por programas *open source*.

Con respecto a la forma de desarrollar software, existen dos opciones: *outsourcing* e *in-house*.

Outsourcing se adapta de mejor manera al modelo de negocios de MINSAL, ya que se entiende que el giro del Ministerio no es producir software, por lo cual resulta mejor externalizar este servicio. Sin embargo, dada la experiencia, se recomienda que sea un equipo de gente preparada en contratos de tecnología dentro de MINSAL, quien controle las licitaciones.

En un largo plazo, la cantidad de piezas de software, sus respectivos mantenedores y los cambios constantes para cumplir con nuevas funcionalidades sugiere la creación de un equipo de desarrollo *in-house*, inmerso en los problemas diarios y consciente de las necesidades de los usuarios. Tener una unidad interna permite controlar la dirección y objetivos de un proyecto mientras este se realiza. Este modelo se ha masificado con la aparición de los *cloud services*.

Francia comenzó en el año 2004 una migración al *open source* (Wired, 2013). Desde entonces, la gendarmería francesa ha impulsado el traspaso de Windows a Ubuntu, creando incluso su propia versión de este sistema operativo de código abierto (GendBuntu). Al año 2017, según la misma gendarmería francesa, son 82.000 computadores los que utilizan este sistema, y ha producido un ahorro en licencias de 7 millones de euros al año, 2 millones correspondientes a licencias de office (ZDNet, 2008). Su paso se ha convertido en una inspiración a nivel internacional para la adopción de estas tecnologías.

A continuación, se presenta un simple ejercicio que ejemplifica el ahorro en licencias de software si MINSAL optara por dejar de utilizar Windows y Microsoft Office, a cambio de un sistema operativo tipo Linux y OpenOffice. Este es solo un ejercicio cuyo objetivo es hacer al lector reflexionar en relación al ahorro que se podría hacer en software, si es que se optara por *open source*. Los valores monetarios son solo referenciales.

Supongamos el documento de Contraloría (2019) que da cuenta de que existen, al menos 36.068 computadores en el MINSAL. Supongamos también un precio promedio anual de licencias¹ de Windows de Enterprise E3 de US\$220 y Microsoft Office 365 de US\$190. Esto daría un costo de licencias por computador de US\$410, lo cual implica un desembolso total de US\$14.787.880. Si bien esa cifra es más alta para una menor cantidad de computadores que la notificada en el caso francés, esto se debe a que estos ocupaban una menor cantidad de licencias que los computadores totales.

¹ Basado en las licitaciones encontradas en las tiendas de Convenio Marco.

Análisis a las Bases de Datos del Sistema de Salud Público

En esta sección se describen las bases de datos a las cuales esta Comisión tuvo acceso, que han sido analizados solo en el formato en el cual se entregaron (no se pudo acceder a los datos en la forma de almacenamiento original).

Aspectos Claves de las Bases de Datos

A continuación se explica brevemente la literatura que rodea las bases de datos. En general, existen dos tipos de bases de datos:

- **Bases de datos relacionales:** son las más comunes. Cada tabla representa a una entidad que se relaciona con otras entidades por medio de tablas de relaciones, que hacen uso de atributos claves (Date, 2000). Cada columna de una tabla representa un atributo (característica) de dicha entidad, y cada fila corresponde a las distintas entradas/valores que la entidad posee.
- **Bases de datos no relacionales:** a diferencia de las bases relacionales, no usan estructuras fijas como tablas, sino que almacenan dinámicamente los datos que llegan bajo una llave, sin un orden particular. El objetivo de estas bases de datos es la rapidez y flexibilidad en agregación y recuperación de datos. Las bases de datos no relacionales son usadas por empresas que almacenan grandes volúmenes de datos dinámicos no estructurados, como los provenientes de redes sociales y, al ser menos usadas que las bases de datos relacionales, existen menos oferta de especializado en ellas.

Considérese también los siguientes conceptos a tratar para la siguiente sección:

- **Atributos:** corresponde a una columna de la tabla. Se le dice también “campo”.
- **Atributo clave:** es la columna que identifica únicamente a una fila dentro de un modelo.
- **Atributos derivados:** son aquellas columnas que se pueden obtener a partir de otra. Por ejemplo, si en una columna se tiene el atributo “Fecha de Nacimiento”, el atributo derivado a partir de ella sería “Edad”.
- **Llave foránea:** dentro de una tabla, es el atributo que hace referencia a una fila única de una tabla externa.
- **Formas normales:** establecen los criterios para evaluar la lógica y construcción de las bases de datos relacionales.

Cabe destacar que las bases de datos que aquí se refieren, se analizan desde el punto de vista de usuario. Es decir, se analizan los datos tal como vienen, pues resulta imposible saber la forma exacta en la cual se almacenan los datos, sin tener un acceso directo al sistema del que provienen las bases de datos.

Descripción de las Bases de Datos

Este estudio tuvo acceso a las siguientes siete bases de datos:

1. **GES:** contiene la lista de espera GES entre los años 2005 y 2017. Los datos se desglosan en dos tablas:
 - **Casos GES:** cada fila de la tabla describe un caso GES. Una persona puede tener más de un caso GES.
 - **Garantías de oportunidad:** cada caso GES da origen a ciertas garantías de oportunidad. De este modo, se entiende una garantía de oportunidad como una atención que hace referencia a un caso GES.
2. **Lista de espera no GES:** contiene la lista de espera no GES entre los años 2014 y 2017. Cada fila de la tabla representa un caso no GES, y una persona puede tener más de un caso no GES asociado.
3. **GRD:** contiene todos los egresos que se realizan en los centros de salud públicos. Detalla para cada hospitalización todos los diagnósticos y procedimientos relacionados, así como también el número GRD de la hospitalización.
4. **Licencias:** comprende todas las licencias solicitadas a Fonasa en el período 2014-2017.
5. **REM (Resúmenes Estadísticos Mensuales):** es un compilado mensual oficial de todas las estadísticas nacionales que se obtienen de los datos de salud. La información contenida en la base de datos REM sirve de apoyo a la gestión central y local y es gestionada por el DEIS. En este estudio, se cuenta con la base de datos REM de los años 2014-2017 (DEIS, 2019).
6. **Tablas quirúrgicas:** describe todas las intervenciones quirúrgicas realizadas durante los años 2014 y 2017 en 22 establecimientos de salud.
7. **UGCQ (Unidad de Gestión Centralizada de Quirófanos):** similar a REM, la base de datos de UGCQ resume el uso de pabellones en hospitales públicos. La base de datos es de responsabilidad de la Subsecretaría de Redes Asistenciales.
8. **SIRH (Sistema de Información de Recursos Humanos):** contiene información con respecto a la gestión de recursos humanos de los hospitales públicos. La base de datos se utiliza principalmente para gestionar el pago de las remuneraciones del personal de salud, pero también sirve como fuente de información para la gestión.
9. **WinSIG-Perc:** detalla el sistema de costeo para el sector de salud. A grandes rasgos, y hablando sólo de la perspectiva en la cual se accedió a los datos, ambas bases de datos contienen datos similares, la diferencia es que Perc es una nueva versión del sistema WinSIG.

La siguiente figura resume las bases de datos de salud a la cual la Comisión tuvo acceso:



Figura 1. Resumen de las Bases de Datos a los cuales tuvo acceso la Comisión.

Indicadores de Calidad de Datos

La producción y almacenamiento de datos son entendidos como un proceso de manufactura (Ballou, Wang, Pazer, & Tayi, 1998). Esto quiere decir que los datos se transforman en información útil para la toma de decisiones y para crear conocimiento. Si los datos no son fidedignos, las decisiones que se toman a partir de los datos no serán efectivas e incluso pueden tener resultados negativos.

En el contexto de que los datos son parte de un proceso de manufactura, se identifican los siguientes tres roles (Strong, Lee, & Wang, 1997):

- **Productores de datos:** quienes “producen” los datos. Están directamente asociados al proceso de producción. En este caso, son los hospitales y centros de atención a pacientes.
- **Almacenadores de datos:** aquellos que proveen los recursos computacionales para almacenar y procesar datos. Se les encargan las tareas de almacenamiento de datos, mantención de sistemas y seguridad. En el contexto del estudio, los almacenadores de datos son las entidades relacionadas a datos dentro de MINSAL.
- **Consumidores de datos:** quienes usan los datos. Los datos se pueden utilizar directamente desde su fuente, o pueden incluir procesos de agregación e integración. Incluye a todos los que hacen uso de los datos de salud, lo cual abarca distintos organismos públicos. Sin embargo, los usuarios directos de los datos son los departamentos de MINSAL a quienes se le reportan datos agregados.

Existen distintos enfoques para evaluar y definir calidad de datos (Wang & Strong, 1996). La literatura, en su mayoría se refiere a un enfoque intuitivo, en el cual se define a la alta calidad de datos como aquella que es útil para los consumidores de datos. Sin embargo, en este estudio no sólo se considera el enfoque intuitivo, sino que también se incluye el enfoque literario de sistemas de información, que comprende distintas dimensiones de calidad.

Para evaluar la calidad de los datos almacenados, se utilizan 11 indicadores que se profundizan a continuación (Strong, Lee, & Wang, 1997). Cabe destacar que para cada indicador se deja a interpretación del lector el mínimo para funcionar. Esto, porque los indicadores requieren un análisis más fino para reflejar el impacto en los pacientes que tendría su mejora cuantitativa.

Nuevamente, se recuerda que el siguiente análisis se hace desde una perspectiva de usuario externo de los datos, y no como productor de los datos o usuario interno.

Accesibilidad

Los datos son fundamentales para obtener estadísticas y realizar análisis cuantitativos, por lo cual mientras menor sea el plazo en el cual se obtengan, más tiempo se dispondrá para estudiarlos y comprenderlos a profundidad. La accesibilidad hace referencia al tiempo que transcurre desde que se solicitan los datos hasta que estos son entregados.

El análisis de accesibilidad indica el nivel de acceso que una entidad externa, distinta al MINSAL, en este caso, tiene a los datos. Esto es particularmente relevante, especialmente si se tiene en consideración que incluso, según lo señalado por el propio MINSAL, dentro del mismo Ministerio existen dificultades para acceder a las bases de datos (Cortés & Quiroz, 2019).

Se utiliza la siguiente ecuación para calcular la accesibilidad:

$$\text{máx} \left\{ 0, 1 - \frac{\text{Tiempo entrega} - \text{Tiempo solicitud}}{\text{Período en los cuales los datos son útiles}} \right\}$$

El resultado de la operación indica la rapidez de respuesta que tiene el organismo al cual se solicitan los datos. Si los datos se entregan después del plazo en que son útiles, el tiempo de respuesta del organismo es de 0%. En caso contrario, si los datos se entregan en el mismo momento en el cual son solicitados, el tiempo de respuesta es del 100%.

En este estudio, el tiempo en que los datos son útiles corresponde a cuatro meses. Se tiene en cuenta que el mandato presidencial que da origen a este estudio es de 12 meses, y se esperaba trabajar con los datos durante 8 meses, por lo cual durante los primeros cuatro meses se esperaba acceder a los datos para realizar los análisis correspondientes. Teniendo en cuenta dicho período, estos son los indicadores:

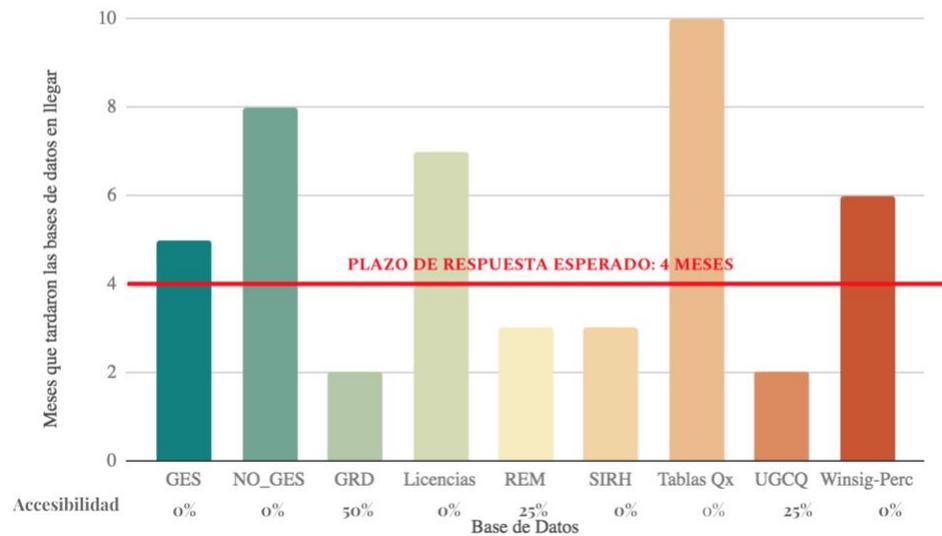


Gráfico 1. Resultados para indicador de accesibilidad. Se establece un plazo de 4 meses máximo para recibir las bases de datos, desde el momento en que se solicitaron. Los resultados indican que 4 de las 9 bases de datos solicitadas se obtuvieron dentro del plazo de utilidad definido.

Los resultados muestran que los datos son de baja accesibilidad. Incluso, en algunos casos los datos se obtuvieron después del plazo de utilidad determinada. La baja accesibilidad de los datos no permite que se pueda generar conocimiento a partir de ellos. Adicionalmente, retrasa el avance de los proyectos para los cuales se requieren los datos.

En este análisis no se incluyeron experiencias contadas por otros investigadores que solicitaron los datos, pero que nunca tuvieron respuesta alguna por parte del Ministerio.

Experiencias internacionales, como las mencionadas en secciones previas de este documento, demuestran que la colaboración de entidades productoras de datos de salud con organizaciones de I+D ayudan a mejorar la atención al paciente. La falta de colaboración entre entidades no da espacio a que se generen investigaciones y nuevos conocimientos. Por lo anterior, se propone democratizar los datos nacionales de salud. Esto implica realizar alianzas con organizaciones de I+D, como universidades, para trabajar con los datos nacionales de salud, siempre teniendo en cuenta de que se trabajen con datos anonimizados.

Con el fin de estar a la vanguardia en cuanto a cyber seguridad, se sugiere revisar el estado del arte en relación a protección y anonimización de datos, como, por ejemplo, privacidad diferencial, la cual es utilizada hoy en día por empresas como Apple.

Pese a lo anterior, se destaca que los proveedores de datos de salud no tienen de por sí ninguna obligación de responder rápidamente a las peticiones de datos, ya que la Ley de

Protección de la Vida Privada (artículo 5), no establece tiempos máximos para responder a las peticiones de transmisión de datos personales. Se sugiere por lo tanto, que esta Ley establezca tiempos máximos para responder a estas peticiones. Estos tiempos máximos se refieren tanto a un tiempo de evaluación (es decir, si se acepta o rechaza la petición) y (en caso de ser positiva la respuesta), el tiempo en que se transmiten los datos.

Todo lo anterior permite concluir que los datos de salud son de baja accesibilidad para aquellos organismos que requieren trabajar con ellos, generando retrasos en proyectos y políticas públicas que dependen de estos. Es más, en algunos casos, los datos se entregan en etapas finales del proyecto, cuando ya no son útiles.

Cantidad de Datos Apropriados

Se refiere a si los datos existentes son suficientes con respecto a los datos esperados.

Este indicador es difícil de cuantificar, ya que para saber si los datos son pocos o muchos, se requiere saber para qué se van a utilizar. Por esto, se establecen dos formas de estimar la suficiencia de estos. La primera consiste en comparar los datos con cifras oficiales, con lo cual la brecha entre ambos indica qué tan apropiada es la cantidad de datos. La otra forma, consiste en obtener una muestra estadísticamente representativa del total y comparar con los datos que se tienen.

El indicador se mide como:

$$\text{máx} \left\{ 0, 1 - \frac{\text{Número de datos apropiados}}{\text{Número de datos existentes}} \right\}$$

En particular, para este indicador sólo se evaluará la base de datos GRD. Esto, porque hasta el año 2017 no era necesario registrar las cirugías ambulatorias que no requieren hospitalización (entiéndase por hospitalización al hecho de pasar más de una noche en el hospital). Para esto, se utiliza como muestra estadísticamente representativa a los registros del Hospital San José de Santiago.

Base de Datos	Porcentaje de cirugías ambulatorias
Tablas quirúrgicas Hospital San José	14,21%
GRD	9,10%
Indicador de datos apropiados	36%

Tabla 1. Resultados para el indicador de cantidad de datos apropiados. Se utiliza una muestra estadísticamente representativa, como los datos del Hospital San José, para comparar los resultados de la base de datos GRD. Los resultados indican que los datos no capturaban antes del 2017 todas las cirugías ambulatorias correspondientes.

Los resultados muestran que la base de datos de hospitalizaciones y procedimientos, hasta el año 2017 no incluía cirugías ambulatorias,² pese a que la naturaleza del proceso de registro es idéntica en ambos casos. Esto redundaría en una mala estimación del uso de recursos de quirófanos, camas y similares, así como también de cualquier estadística general con respecto a hospitalizaciones.

El caso mencionado es muy particular y atinente solo a una base de datos, otras bases no tienen casos de esta magnitud.

Se sugiere realizar un catastro de las excepciones de registro que se efectúan, e incluirlas en su correspondiente base de datos. Esto debe ir acompañado de una revisión de todos los procesos que no están siendo correctamente reflejados en la base de datos.

Concisión

Se define como la capacidad de la base de datos de representar de forma exacta y breve su contenido. Una base de datos concisa es auto contenida, es decir, su contenido es suficiente para comprender la realidad que captura.

La concisión se define considerando la relevancia de las columnas, lo cual indica la importancia de los campos que son parte de la base de los datos. En específico, se evalúan los siguientes aspectos:

- Número de columnas que no entregan información, es decir, columnas que, si se restan, no modifican la significancia de las filas.
- Número de columnas que no están en primera forma normal (Codd, 1989). Las formas normales permiten evaluar la lógica con la cual están construidas las bases de datos. En específico, la primera forma normal dice que las tablas deben estar libres de repeticiones en cuanto a su contenido.
- Número de columnas que no están en tercera forma normal (Codd, 1989). La tercera forma normal indica que ningún atributo debe depender de parte parcial de la clave.
- Número de columnas que contienen atributos derivados (por ejemplo, se tiene una columna que indica la fecha de nacimiento del paciente y otra que indica la edad del paciente).
- Número de campos cuyo contenido es redundante con los datos entregados por otra tabla.

Una base de datos concisa ahorra tener que lidiar con datos que no entregan valor, y sobre todo, se reducen los costos computacionales relacionados a un mayor tamaño de la base de datos. Entre estos costos, se incluye un mayor uso del espacio en disco, un menor tiempo de respuesta (ralentización de las consultas) y mayor complejidad de cómputo de las consultas.

² Se desconoce el motivo de esto.

El indicador se define como:

$$\text{máx} \left\{ 0, 1 - \frac{\text{Columnas que violan la concisión}}{\text{Total de columnas}} \right\}$$

El indicador señala el porcentaje de concisión que tiene la base de datos. Una base de datos que es estrictamente concisa tiene un 100% de concisión, mientras que una en la cual los atributos no entregan mayor información, está más cercana al 0% de concisión. Los resultados para este indicador corresponden a:

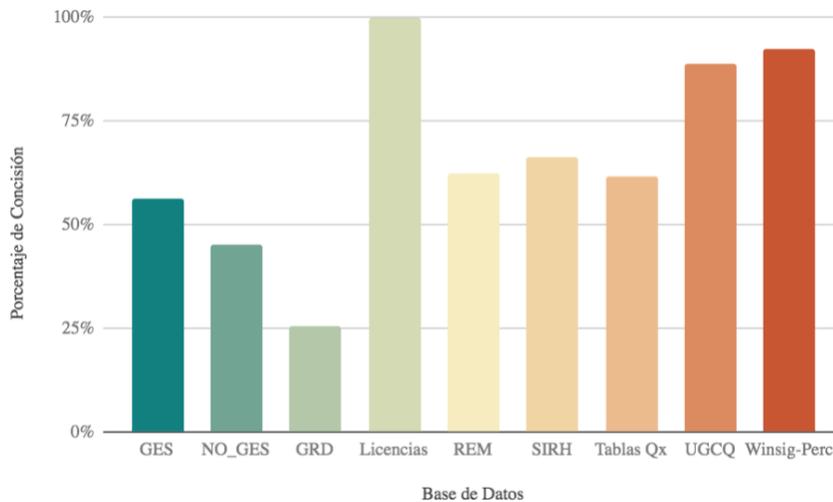


Gráfico 2. Resultado para el indicador de concisión. La base de datos Licencias es la única que tiene un 100% de concisión, mientras que GRD es la base de datos que tiene mayor cantidad de columnas que no agregan valor.

En el análisis se encuentran atributos derivados no actualizados en las bases de datos. Por ejemplo, en la base de datos GES existe una columna que muestra el estado del caso en el mes anterior, pero no queda claro si es que se refiere al estado del mes anterior a la atención o no, lo que evidencia un desconocimiento del funcionamiento de datos de este tipo. Esto, porque si un dato depende de otro, entonces debiese actualizarse automáticamente junto a él. Es un malgasto de recursos computacionales almacenar información actual junto a la desactualizada.

También, existen datos almacenados que replican información de poco valor. En otras palabras, hay datos derivados duplicados, cuya existencia se debe a que son útiles sólo para un usuario específico. Por ejemplo, en la base de datos GES hay 8 columnas que dan cuenta de la edad del paciente al momento de crear el caso y de ser atendido. Adicionalmente, ya se cuenta con las columnas de fecha de nacimiento del paciente y las fechas correspondientes de atención, con lo cual la se puede deducir la edad. En otras palabras, 8 columnas de la base de datos GES sobran ocupando recursos computacionales

Se concluye que, en general, las bases de datos no son concisas y como consecuencia, tampoco son escalables. La falta de escalabilidad implica que el sistema no responderá de forma adecuada ante un aumento en la cantidad de pacientes históricos.

Para resolver la concisión, se propone utilizar un modelo de bases de datos entidad-relación, el cual comprenda todas las bases de datos y que cumpla con las formas normales. El modelo de base de datos entidad-relación resuelve el problema de escalabilidad, y según la literatura, es el más validado para datos estructurados como los que se presentan. Además, se sugiere usar sistemas de bases de datos que actualicen oportunamente los datos derivados.

Si bien las alternativas *NoSQL* (bases de datos no relacionales) podrían tener un mejor desempeño y escalabilidad para la velocidad de las consultas, estas requieren una alta cantidad de tiempo para indexación y tiene problemas en mantener las propiedades ACID (atomicidad, consistencia, aislamiento y durabilidad), necesarias para una base de datos de esta importancia. Es decir, en este ámbito, las bases *NoSQL* son favorables para analizar datos y consultas de población, pero no para acceder a un valor específico (como una ficha clínica) del mismo modo que lo haría un sistema *SQL*. La literatura sugiere un consenso entre ambas alternativas, manteniendo una copia que no se actualiza en tiempo real para el análisis global (Miranda, y otros, 2016).

Consistencia

Mide la coherencia en las bases de datos. Una base de datos consistente no tiene atributos que se contradigan entre sí. Si la base de datos no es consistente, será poco creíble.

Para medir los casos inconsistentes, se consideran el número de columnas que tienen datos contradictorios entre sí. La fórmula del indicador es:

$$\text{máx} \left\{ 0, 1 - \frac{\text{Columnas que violan la consistencia}}{\text{Total de columnas}} \right\}$$

En general, las bases de datos cumplen con el criterio de consistencia. Sin embargo, vale la pena mencionar que la base de datos GRD, donde está la columna GES que indica si la patología tratada es o no GES. Adicionalmente existe la columna Programa Social, la cual informa, entre otras categorías, si la patología tratada es o no GES. Los totales de estas bases de datos no calzan.

En la base de datos GES, existen tres columnas que hacen referencia al hecho de que si se confirmó el Informe de Proceso Diagnóstico. Estas columnas corresponden a: IPD_confirmado, IPD_descartado y caso_descartado_por_IPD. Los números totales de estas 3 columnas no calzan, es decir, existen diferencias entre el total de casos con IPD confirmado y descartado entre las filas.

Se observa un ejemplo de inconsistencia en la base de datos GES. Existen tres columnas que hacen referencia a si se confirmó el Informe de Proceso Diagnóstico. Estas columnas corresponden a: IPD_confirmado, IPD_descartado y caso_descartado_por_IPD. Los números totales de estas 3 columnas no calzan, es decir, existen diferencias entre el total de casos con IPD confirmado y descartado entre las filas.

Lo anterior indica que las bases de datos son poco creíbles y que además se están malgastando recursos computacionales.

Como solución, se propone hacer una revisión del historial de los pacientes para resolver las contradicciones en los datos. Para ello, se sugiere seleccionar una de las columnas que se contradicen para almacenar la información correctamente, y eliminar las otras columnas.

Compleitud

Consiste en evaluar qué tan poblada está la base de datos. Una base de datos poblada es aquella en la cual no hay valores vacíos. La completitud se define como:

- Valores “*missing*”, que son el número de campos vacíos.
- Existencia de categorías definidas que no se encuentran en los datos. Por ejemplo, se definen 5 categorías posibles para un campo, pero sólo se encuentran 4 de ellas. También considera el ejemplo contrario, en que se definen 5 categorías posibles para un campo, pero se encuentran 6 posibles valores.
- Temporalidad: se refiere a aquellos casos en que se debe cumplir una secuencia de pasos, en la cual uno de los pasos no se cumple.

Si bien utilizar valores *null* no es necesariamente malo (contrario a ciertos autores de bases de datos), su mal uso es frecuente y ocurrente dentro de las bases de datos (Date, 2004; Pascal, 2012). Estos pueden afectar seriamente los resultados de las consultas, y por lo tanto no debe usarse un valor *null* en información que debería conocerse, ejemplos de estos son la confirmación diagnóstica en la lista de espera No GES, o la descripción de la prestación FONASA en la base de datos GES (Ramakrishnan & G, 2000).

El indicador se define como:

$$\max \left\{ 0, 1 - \frac{\text{Número de campos vacíos}}{\text{Total de campos}} \right\}$$

Los resultados para este indicador se presentan a continuación:

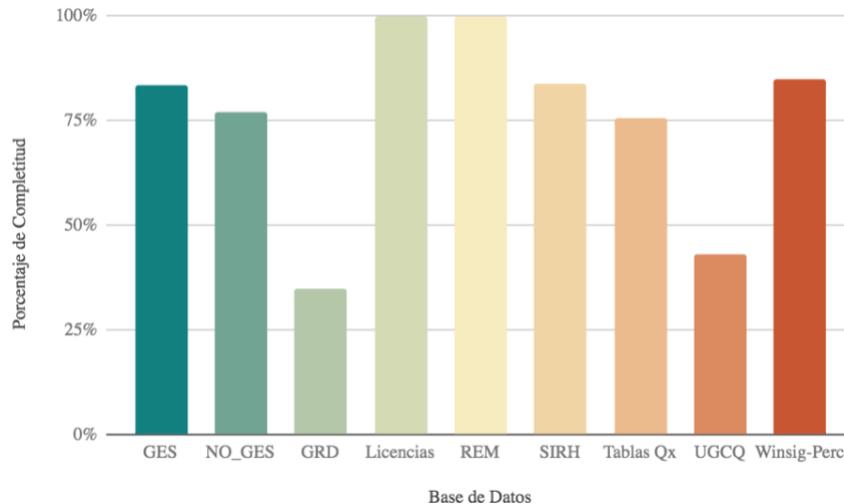


Gráfico 3. Resultados para el indicador de completitud. Los resultados muestran que la base de datos de GRD es aquella más incompleta. Se destaca que en la base de datos GRD se cumplen todas las incompletitudes de las bases de datos, inclusive las incompletitudes de categorías. Por ejemplo, para el campo sexo del paciente, se encuentran 34 formas diferentes en que se escriben los sexos del paciente, mientras que el instructivo de GRD define . También, el instructivo de GRD define cuatro causales de salida, pero en los resultados se encuentran 15 opciones.

En la mayoría de las bases de datos existen datos faltantes: *missing values*. Esto dificulta cualquier obtención de estadísticas e intención de reconstruir el historial del paciente.

Como solución para los datos ya almacenados, se propone reemplazar los *missing values* buscando los datos en otras bases de datos. Por ejemplo, en el caso de que en una base de datos falte la fecha de nacimiento del paciente, se recomienda buscarlo en otras bases.

A futuro, se sugiere que los datos capturen toda la realidad. Esto implica guardar todo dato que el paciente entregue y los metadatos de la interacción entre este y el sistema. Además, es necesario poder obtener información de otros servicios para entender de mejor forma las causas de ciertos eventos que muestran los datos.

La alta cantidad de valores vacíos imposibilita reconstruir el historial médico de un paciente. Esto puede incluso llegar a afectar el tratamiento del paciente, al no poder reconstruir su historial médico.

Se propone comenzar a indexar de manera correcta todo lo que ya se registra (pues el costo asociado es mínimo) junto a datos que sean parte del mismo evento. De ahí en adelante es necesario contar con formas de registro digital que permitan no solo una estandarización de lo ya registrado, sino que permitan registrar más información sin poner una carga sobre el personal médico.

Si bien, desde el punto de vista matemático, es deseable tener un conjunto mínimo de datos a almacenar, en la práctica, el costo adicional de almacenar un atributo extra de un hecho que ya se registra, es nulo. La literatura de *Data Science* y la práctica de grandes empresas que manejan datos (como Facebook, Apple, Google), señalan que mientras más datos se posee, más información se puede extraer para mejorar el sistema, pues las herramientas de hoy permiten analizar masivas cantidades de datos.

Por último, cabe destacar que la literatura sobre conjunto mínimo de datos varía según práctica y según país. Por ejemplo, las prácticas de enfermería muestran datos subjetivos y objetivos, (como el peso del paciente o si el paciente trabaja en algo que le dé sentido a su vida, o incluso la posición que asume para dormir mientras está hospitalizado). Por otro lado, países como Estados Unidos y Canadá recogen tanto valores de exámenes como información financiera y personal. La literatura ha determinado que los estándares utilizados internacionalmente no son suficientes para ser concluyentes. Dado lo anterior, la alternativa más viable es apuntar hacia el futuro, utilizando la estrategia que hoy en día usan las grandes empresas sobre los datos que analizan. Esto implica guardar todo dato que el paciente entregue y los metadatos de la interacción entre este y el sistema. Adicionalmente, es necesario poder obtener información de otros servicios para entender de mejor forma las causas de ciertos eventos que muestran los datos.

Credibilidad

La credibilidad indica qué tan ciertos son los datos. Para tener certeza de los datos almacenados, se comparan los valores de las tuplas con estándares del medio de salud. Este indicador considera las tuplas espurias, es decir, aquellas que no tienen valor.

Puesto que un análisis de credibilidad requiere un conocimiento médico técnico que escapa del alcance de este proyecto, sólo se menciona el caso de la base de datos GES, en la cual existen tuplas cuya información es errónea, para la cual se da un ejemplo a continuación:

Para **8.788** casos GES de **Cáncer de testículos**

La descripción del estado del caso corresponde a **Tratamiento Fisura Palatina**

Figura 2. Ejemplo de contradicciones técnicas en la base de datos GES.

Estas tuplas erróneas llevan a una poca credibilidad de las bases de datos.

Facilidad de Manipulación

La facilidad de manipulación hace directa referencia a la interoperabilidad de las bases de datos. En otras palabras, se refiere a qué tan simple es cruzar bases de datos, es decir, encontrar datos dentro de otra base de datos.

Para este indicador en particular, se busca responder a preguntas del tipo ¿qué tan fácil es encontrar a un paciente de la lista de espera en la base de datos GRD? Puesto que la Comisión tiene acceso a los datos anonimizados de los pacientes, se solicitó ayuda a MINSAL para cruzar las bases de datos. Sin embargo, con los datos no sensibles, como por ejemplo el código identificador del establecimiento de salud, se encontraron inconsistencias en la forma de escribir el código.

A continuación se presenta un ejemplo en el cual se intentaron cruzar los identificadores de los establecimientos de la base de datos GES (perteneciente a FONASA) con la lista de establecimientos oficiales publicada por DEIS. Los resultados son los siguientes:

170 establecimientos que se encuentran en la base de datos GES no aparecen en las bases de establecimientos del DEIS.

- De estos 170, se pudieron rescatar 96 establecimientos, buscando la equivalencia de nombres:
 - Sólo se asignó el establecimiento si y sólo si el tipo de establecimiento era el mismo. Ejemplo: "Centro de Salud Comunitaria Familiar Santa Bárbara" por "CECOSF Santa Bárbara".
 - No se asignaron establecimientos que no eran del mismo tipo, pues tomar supuestos podía inducir a errores. Por ejemplo, un establecimiento no encontrado es el de "Pedegua, Posta". Al buscar algún alcance de nombres en la base de datos de establecimientos, lo más similar es el "Centro Comunitario de Salud Familiar Pedegua". Puesto que no hay forma de asegurar de que la posta se haya transformado en CECOSF, es que no se asignó el código de una a otra.
 - En algunos casos, no había equivalencia de nombres, pero al realizar una búsqueda más profunda, se encontraron otros tipos de relaciones. Por ejemplo, existía el "centro de diagnóstico y tratamiento Eloísa Díaz", el cual al investigar de forma más profunda se encuentra que es parte del complejo hospitalario del San José.

De los 170 establecimientos, para 59 de ellos no había ninguna ocurrencia. Cada uno de estos establecimientos se buscó en internet para validar su existencia, y se encontró que son establecimientos de atención primaria. Se concluye además:

- La suma de los 59 establecimientos sin datos, más los 96 establecimientos encontrados, da un total de 155 establecimientos, lo cual genera una diferencia de 15 establecimientos para completar los 170 en total. ¿dónde están los 15 restantes para llegar a 170?
- La respuesta es que 12 de estos establecimientos tenían id "0", pero distintos nombres. Por ejemplo, al buscar el código 0 aparecen las siguientes descripciones "Araucanía norte", "MINSAL". Esto sucede 12 veces.
- Los 3 restantes son establecimientos cuya descripción es demasiado vaga como para poder recuperarlos, y la información de las demás columnas, como la comuna, no otorgó suficientes pistas para descubrir qué establecimiento es. Por ejemplo: "Establecimiento Prais".

Lo anterior deja en evidencia la dificultad para interoperar bases de datos, pues tan solo realizar un cruce que permita identificar establecimientos en las distintas bases de datos involucra un alto costo operacional. Las bases de datos no están pensadas como una base de datos en sí, sino que como grupos de tablas aisladas. Las bases de datos se alimentan de distintos sistemas de información, y hasta el momento no se está realizando una estandarización de datos centralizada que permita que las tablas conversen entre sí.

Así, se proponen las siguientes acciones para solucionar el tema: a corto plazo, se recomienda realizar una limpieza de datos para que estos sean interoperables. Esto quiere decir, definir los atributos primarios y foráneos, para luego estandarizarlos. Por ejemplo, utilizar siempre el mismo código para referirse a un establecimiento de salud. A largo plazo, se sugiere integrar todas las bases de datos en un modelo relacional.

Interpretabilidad

Indica qué tan entendibles son los datos. Se considera que la base de datos es entendible si:

- El nombre de las columnas, dentro de lo posible, es autocontenido. El nombre de cada columna, en lo posible, debe estar apegado a las convenciones y buenas prácticas de bases de datos.
- Los valores de las tuplas son autocontenidos también. Los valores que llenan la base de datos deben ser fáciles de comprender.

La interpretabilidad se calcula como:

$$\text{máx} \left\{ 0, 1 - \frac{\text{Número de columnas difíciles de interpretar}}{\text{Total de columnas}} \right\}$$

Los resultados son:

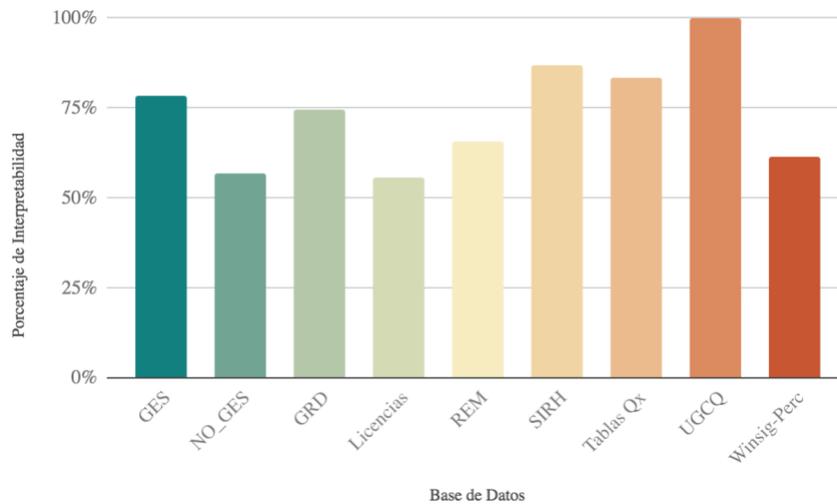


Gráfico 4. Resultados para el indicador de interpretabilidad. Los resultados muestran que la base de datos de Licencias y lista de espera NO GES son las que peor desempeño en cuanto a interpretar su contenido.

Dentro de los resultados más destacables, destacan los siguientes nombres de las columnas en GRD: v10, v17, v60, de las cuales se desconoce el significado de su contenido.

Las bases de datos son difíciles de interpretar, principalmente porque los nombres de los atributos no cumplen con las buenas prácticas de computación. Esto implica un alta barrera de entrada para entender el contenido de las bases de datos y sus tablas.

Se propone implementar buenas prácticas internacionales de bases de datos y computación para incrementar la interpretabilidad.

Libre de error

Esta dimensión hace referencia a los errores de contenido de las bases de datos. Comprende las siguientes dimensiones:

- Datos mal escritos.
- Errores de tipo de datos. Cada atributo debe especificar el tipo de datos que almacena, como por ejemplo *string*, *integer*, *boolean* (FOLDOC, 2003). Este tipo de error se refiere a cuando se almacenan datos que no corresponden al dominio (por ejemplo, se define el tipo de un atributo como *integer* -números-, pero en realidad se almacena texto).

Este indicador comprende las otras categorías. Por ejemplo, el error en tipo de almacenamiento de datos se incluye dentro del indicador de completitud. Por ello, se destaca en específico errores de texto.

Se menciona el caso de la base de datos SIRH, para la cual el campo de descripción del cargo del profesional de la salud está en texto libre. El siguiente gráfico representa las variaciones de texto que existen en esta base de datos para identificar a los técnicos en enfermería (TENS):



*TENS: Técnico en Enfermería Nivel Superior

Gráfico 5. Descripción de funciones ejercidas por profesionales TENS (Técnicos en Enfermería Nivel Superior). El gráfico representa la dificultad que existe en la base de datos SIRH para identificar a los TENS. Un 4% de los TENS tienen una descripción de función no atinente a su labor.

Se recalca también la base de datos NO GES. En la cual existen texto en formato libre para los diagnósticos. Para los siguientes ejemplos, se considera la columna “confirmación diagnóstica”.

Ejemplos y frecuencias de textos que no informan diagnósticos

22.685

“0”

15.377

“No registra diagnóstico confirmado”

6.599

“OTRO”

5.424

“Consulta no Especificada”

5.402

“No Informado”

3.869

“1”

2.062

“SI”

1.866

“...”

Ejemplos y frecuencias de textos en el cual existen diferentes formas para escribir el diagnóstico “vicios de refracción”



Ejemplos y frecuencias de textos en los cuales se ha escrito texto que no corresponde al dominio del campo. Esto quiere decir, que el texto no tiene relación alguna con el diagnóstico del paciente*.



*Por motivos de protección de datos personales, se han cubierto los datos del profesional al cual el texto hace referencia.

Ejemplos y frecuencias de textos en el cual el diagnóstico en texto libre no es específico



Figura 3. Ejemplos de imprecisiones en el uso de textos libres para describir el diagnóstico médico en la base de datos no GES.

Las variaciones, abreviaciones y alteraciones que poseen valores que debieran representar una misma categoría, hacen que sea imposible de normalizar y obtener estadísticas concluyentes a partir de estos datos.

Asimismo, se destaca que no existen protocolos de transmisión y almacenamiento de datos. Sin embargo, existen manuales de buenas prácticas desde MINSAL con respecto al formato en que se almacenan los datos, los cuales, según nuestro análisis de calidad de datos, no han surtido efecto.

Como solución, para los datos que ya están mal escritos, se recomienda utilizar técnicas de minería de texto para limpiar los datos. Con respecto a los datos que están por ser incluidos a futuros, se sugiere modificar los campos de texto libre por selección de opciones y/o verificación automática de formato.

Se sugiere también eliminar del proceso de captura de datos a los digitadores. Es decir, no debe pasar de un papel por medio de un digitador a la base de datos. La captura de datos debiese ser inmediata y en formato digital. Además, se sugiere reforzar al personal médico el uso de herramientas digitales, con énfasis en la codificación de enfermedades en términos de CIE-10, con fines estadísticos. Si bien CIE-10 es útil para la gestión estadística, tiene ciertas deficiencias para la gestión clínica, como por ejemplo el hecho de que no son capaces de almacenar los detalles entregados por los médicos de forma que les sea útil más tarde (Mauro, 2019). Por ello, se sugiere también explorar la herramienta SNOMED-CT, que propone alternativas de solución al problema.

Temporalidad

Consiste en determinar qué tan actualizados están los datos para el propósito que se requieren. En cuanto a políticas públicas se refiere, lo ideal es que los datos estén actualizados al tiempo real, de modo tal de que sean un fiel reflejo de la realidad (Giannone, Reichlin, & Small, 2008). Una base de datos actualizada a tiempo real es una base de datos en la cual los datos son válidos (Stankovic, Son, & Hansson, 1999).

En específico para este indicador, se toma en cuenta las listas de espera tanto GES como no GES.

Para el caso de la lista de espera GES, se consideran dos columnas: la primera de cuenta de la fecha en la cual se otorga una garantía de oportunidad específica para un caso, mientras que la segunda columna detalla la fecha en la cual se crea dicho caso GES. Se entiende que la creación del caso GES es previa al otorgamiento de las garantías de oportunidad que derivan de dicho caso. Los resultados de este análisis se presentan en la siguiente figura:



Figura 4. Análisis de temporalidad para la fecha de creación del caso GES y la fecha de otorgamiento de garantías de dicho caso. La fecha de creación de un caso GES debiese ser menor a la fecha en la cual se otorgan las garantías de oportunidad que se derivan de ese caso, lo cual se cumple efectivamente para el 10% de las filas de la base de datos GES.

El análisis anterior es un ejemplo de que existen inconsistencias en cuanto a la temporalidad de la lista de espera GES.

Para la lista de espera no GES, se solicitó ayuda a MINSAL para contar la cantidad de casos GES en dos cortes distintos. En específico, se hace el siguiente ejercicio:

- Se comparan dos versiones de la lista de espera NO GES. Una versión fue extraída en septiembre del año 2018, mientras que la otra versión corresponde a abril del 2019.
- Se definió un período de tiempo de análisis entre 01/01/2014 y 31/12/2017 y una cantidad de 1.454 establecimientos.
- Los resultados para los mismos establecimientos y el mismo tiempo muestra los siguientes resultados:

	Versión Septiembre 2018	Versión Abril 2019	Diferencia
Casos	9.140.045	9.301.694	161.694

Tabla 2. Ejemplo de aplicación del indicador de temporalidad. Se analiza la lista de espera NO GES para el período 01/01/2014 – 31/12/2017, pero para una versión a septiembre del año 2018 y otra versión a abril del año 2019. Los resultados indican que para un mismo periodo, con el pasar de los meses se añadieron nuevos casos a la lista NO GES.

Esto da evidencias de que existe una descoordinación entre los hospitales productores de datos, y los almacenadores de datos (MINSAL) en cuanto a la actualización de las bases. En específico, puesto que cada hospital controla la información que envía a MINSAL, existen diferencias en la lista de espera no GES que mantienen los hospitales versus la que mantiene MINSAL. Esto implica que la lista NO GES no estima correctamente la cantidad de pacientes en espera.

Según fuentes internas de MINSAL, la desactualización de la lista de espera no GES se debe a que no existe un sistema centralizado con la información de la lista, sino que simplemente se almacenan los datos en distintas planillas de Excel.

Se propone que MINSAL debe estar a cargo de proveer el sistema de información a los hospitales para registrar la lista de espera, lo que implica que exista un único sistema de información para el registro oficial de lista de espera, tanto GES como no GES.

Para asegurarse de que todos los casos se actualicen a tiempo en el sistema, se sugiere implementar incentivos de reducción de lista de espera, así como también dar al paciente el control sobre su registro. De este modo, se asegura transparencia del proceso de lista de espera para todos los actores.

Seguridad

Los datos por sí mismos no permiten establecer un indicador de su seguridad, ya que este debe ser basado en el sistema que los almacena. Solo se puede establecer la seguridad con base en los casos históricos.

Se mencionan algunos casos en los cuales han existido fallas informáticas. En el año 2016, se canceló la continuación del proyecto SIDRA II por no cumplir con las nuevas normativas de ciberseguridad, ya que incluían carpetas compartidas de información, en las cuales eventualmente podría almacenarse datos sensibles. En el mes de julio del 2019, la empresa ECM, que suministra una plataforma de imagenología al Servicio de Salud Metropolitano Sur Oriente, sufrió un ataque informático por lo cual tuvieron que temporalmente suspender sus servicios para resguardar los datos; cabe recalcar que un ataque de este tipo no solo pudo exponer los datos, sino que también eliminarlos completamente de los sistemas (BioBio Chile, 2019). En específico, se revisa a continuación el caso de la filtración de datos del año 2016.

Se propone establecer protocolos de respuesta ante incidentes informáticos (fugas de datos, reportes de errores o problemas de seguridad). Con respecto a los problemas de seguridad y reporte de errores en específico, estos deben ser arreglados en el menor tiempo posible luego de ser reportados. Es importante hacer uso de información reportada por *grey hat hackers* para mejorar los sistemas.

El año 2016 se dio a conocer una falla en el sistema informático de la red del MINSAL que permitió a todo usuario con acceso a la red el poder extraer los datos personales médicos de los pacientes sumando un total de 3 millones de archivos. Pese a que la falla se dio a conocer con 10 meses de anticipación, no se actuó a favor de la protección de los datos por medio de un protocolo pese a que los datos son de carácter sensible (CIPER, 2016). Por un lado, se mostró la falta de uso de herramientas básicas de seguridad tales como encriptación y manejo de permisos. En otras palabras, la barrera de seguridad era tan baja que permitía que cualquier persona dentro de la red tuviera

acceso a los datos sin necesidad de altos conocimientos en computación.

Sumado a esto, gran parte de la información sensible de salud no se guarda en un servidor protegido sino en artefactos personales de quienes trabajan con los datos. Si bien se desconoce la seguridad con lo que los datos son tratados en estos casos, situaciones como esta aumentan el riesgo de fallas de seguridad y robo de datos.

La norma técnica encargada de regular la seguridad y confidencialidad de los documentos electrónicos usados por la Administración del Estado, dictada en el año 2004, no es más que un manual de usuario destinado a los jefes de los servicios públicos, con el propósito de que estos instruyan a sus subordinados en buenas prácticas relativas a seguridad de los sistemas informáticos. Estas instrucciones y deberes son, en términos de la norma, conocidos como nivel básico de seguridad para el documento electrónico y están desactualizadas de las tecnologías de hoy en día.

La norma técnica mencionada anteriormente, no guarda relación alguna con las amenazas de seguridad a las que, como mínimo, un servidor debería estar preparado. En otras palabras, la norma señala simplemente herramientas de seguridad para navegación web en computadores personales. Por esto, se propone solicitar la actualización y reformulación de dicha norma técnica.

Se sugiere contratar personal formado específicamente en seguridad informática junto con constantes capacitaciones al personal encargado. La velocidad con que aparecen nuevos

ataques es tal, que los expertos en esta área deben estar constantemente a la vanguardia en el tema.

En resumen, los sistemas de información del MINSAL han sufrido ataques informáticos, debido a distintas vulnerabilidades detectadas en los sistemas, las cuales incluso pueden llevar a una filtración de datos sensibles personales de los usuarios del sistema de salud.

Se destaca que no hay penalización para una mala manipulación de datos sensibles. La Ley de Protección de Vida Privada no tipifica explícitamente como delito la filtración de datos sensibles.

Por ello, se propone imitar la regulación de protección de datos (GDPR por sus siglas en inglés) de la Unión Europea (Brandom, 2018). La regulación multa a toda aquella organización que filtre datos personales con altas sumas monetarias, donde el mayor castigo corresponde a una multa del 4% de las ganancias de la compañía o 20 millones de dólares (se selecciona al mayor monto entre ambos).

Conclusión

Este documento realiza una revisión y análisis de las bases de datos del sector de salud. En resumen, se encuentran deficiencias con respecto a la calidad de los datos, principalmente en cuanto a las bases de datos que almacenan información de pacientes. Aun así, con un debido procesamiento, es posible realizar análisis básicos sobre esta información.

La deficiente calidad de datos trae múltiples consecuencias, ya que la calidad de atención de los pacientes se ve directamente afectada, ¿Cómo puede un paciente estar seguro de que lo van a atender a tiempo si sus datos no están bien escritos? Desde información como datos personales (número de teléfono, dirección, etc.) podrían estar mal digitados, imposibilitando el futuro contacto, hasta incluso que se registre una cirugía diferente a la que se realizó. ¿Cómo podría esto reflejarse en los estados financieros y contables de dicho hospital?

Se requiere también hacer un catastro del software y hardware existente en salud. Ante la escasez de recursos monetarios, es necesario planificar adecuadamente cómo se invierten los recursos estatales en sistemas de información. Además, si se van a invertir altos montos de dinero en tecnología, es importante hacer seguimiento de la rentabilidad que generan dichos sistemas. El proyecto SIDRA lleva 11 años en funcionamiento y al día de hoy aún no existe la Ficha Clínica a en todos los establecimientos a nivel nacional.

Si bien el panorama actual en cuanto a calidad de datos es sombrío, existen espacios para mejora. En relación a los datos ya almacenados, es importante realizar una limpieza y revisión de datos exhaustiva, de modo tal de arreglar todos los errores encontrados. En cuanto al futuro, se sugiere replantear la forma en la cual se almacenan datos y cómo hacer un sistema de bases de datos robusto.

En Chile existen buenos ejemplos de cómo la tecnología y los datos son un apoyo para la calidad de atención de los pacientes, como es el caso de SISMAULE o el proyecto Almohadita - Ámbar.

Por último, es importante mantener la seguridad y privacidad de los datos. Esto no sólo para los datos del sector de salud, sino que para todos los datos personales que circulen tanto a nivel de empresa como estatal. Actualmente, las leyes hoy en día no son suficientes para proteger los datos y cada día aumenta la frecuencia con la cual se realizan ataques informáticos a organizaciones de todo tipo. Es urgente que se regularice la situación para proteger a las personas.

Trabajo Futuro

Si bien este trabajo presenta numerosa evidencia de problemas en las bases de datos y sus posibles soluciones, existen variadas direcciones en las cuales puede mejorar.

En primer lugar, dado que este estudio se realiza desde un punto de vista externo, es necesario ejecutar el mismo estudio en las bases de datos originales, junto con información que no se posee sobre sistemas internos tales como la seguridad de los mismos.

Además, se sugiere realizar análisis similares con bases de datos de países que tengan buenas prácticas de tecnología, para obtener puntos de referencia hacia los cuales apuntar.

Por otro lado, existen indicadores como el “Libre de error” cuyos resultados son extremadamente dependientes de un humano, encontrar palabras mal escritas para términos de salud de forma automática para ser comparados con un indicador requiere de herramientas que por temas de tiempo y recursos no pudieron ser usadas.

Siguiendo en esta línea, se busca en un futuro reevaluar los indicadores con mayores recursos y acceso interno con el objetivo de obtener resultados más precisos.

Bibliografía

- Contraloría. (2019). *Informe Final 187-2109 Subsecretaría de Redes Asistenciales sobre Auditoría a las Adquisiciones de Servicios Informáticos Efectuadas por la Subsecretaría Junio 2019*. Santiago: I Contraloría Regional Metropolitana de Santiago.
- Centro de Políticas Públicas UC. (2017). *La espera en el sistema de salud chileno: una oportunidad para poner a las personas al centro*. Pontificia Universidad Católica de Chile, Santiago.
- Office of the National Coordinator for Health Information Technology. (23 de 08 de 2019). *Official Website of The Office of the National Coordinator for Health Information Technology (ONC)*. Obtenido de healthit.gov: <https://www.healthit.gov/faq/what-electronic-health-record-ehr>
- TechTarget. (08 de 2017). *SearchHealthIT*. Obtenido de SearchHealthIT: <https://searchhealthit.techtarget.com/definition/electronic-health-record-EHR>
- Kalra, D. (2016). Electronic Health Record Standards. *IMLA Yearbook of Medical Informatics*.
- Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., & Taylor, R. (2005). Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs. *Health Affairs*, 24(5), 1103-1117.
- King, J., Patel, V., Jamoom, E., & Furukawa, M. (2014). Clinical Benefits of Electronic Health Record Use: National Findings. *Health Information Technology*, 49(1), 392-404.
- Fisher, B., Bhavnani, V., & Winfield, M. (2009). How Patients Use Access to their Full Health Records: A Qualitative Study of Patients in General Practice. *Journal of the Royal Society of Medicine*, 102(12), 539-544.
- Blumenthal, D. (2009). Stimulating the Adoption of Health Information Technology. *New England Journal of Medicine*, 360(15), 1477-1479.
- Wang, S., Middleton, B., Prosser, L., Bardon, C., Spurr, C., Carchidi, P., & Kuperman, G. (2003). A Cost-Benefit Analysis of Electronic Medical Records in Primary Care. *The American Journal of Medicine*, 114(5).
- FALMED. (08 de 2019). *FALMED*. Obtenido de Ficha Clínica Electrónica: <http://www.falmed.cl/prevencion-y-educacion/ficha-clinica-electronica.aspx>
- Varas, J., Hering, E., & Demetrio, A. (2010). Ficha Clínica: Composición, Manejo y Registros Clínicos. *Revista Obstétrica y Ginecológica del Hospital Santiago Oriente Dr. Luis Tisné*, 5(1), 57-61.
- Decreto N°41. (2012). Decreto N°41 que aprueba Reglamento sobre Fichas Clínicas, Subsecretaría de Redes Asistenciales, Ministerio de Salud. Santiago.
- Godoy, J., & Barraza, J. (2018). La Ficha Clínica Mirada desde la Legislación Chilena Actual. *Acta Bioethica*, 24(2).
- Johns Hopkins. (03 de 2016). Command Center to Improve Patient Flow. *Insight*.
- Roth, M. (07 de 06 de 2018). 4 Insights to Build a Better Command Center. *Health Leaders*.
- Johns Hopkins. (10 de 2016). *The Johns Hopkins Hospital Launches Capacity Command Center to Enhance Hospital Operations*. Obtenido de Johns Hopkins Medicine:

- https://www.hopkinsmedicine.org/news/media/releases/the_johns_hopkins_hospital_launches_capacity_command_center_to_enhance_hospital_operations
- NHS. (20 de 06 de 2019). *About NHS Digital*. Obtenido de NHS Digital: <https://digital.nhs.uk/about-nhs-digital>
- NHS. (2014). *Personalised Health and Care 2020*.
- NHS. (07 de 2019). *National Indicator Library*. Obtenido de NHS Digital: <https://digital.nhs.uk/data-and-information/national-indicator-library>
- NHS. (2018). Information and Technology for Better Health and Care. *Data, Insights and Statistics*, 1(1).
- NHS. (09 de 2018). *Open Exeter*. Obtenido de NHS Digital: <https://digital.nhs.uk/services/nhais/open-exeter>
- NHS. (07 de 2019). *Keeping Patient Data Safe*. Obtenido de NHS Digital: <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe>
- Framingham Heart Study. (07 de 2019). *Bienvenido al Estudio del Corazón de Framingham!* Obtenido de Framingham Heart Study: <https://www.framinghamheartstudy.org/fhs-about/about-fhs-en-espanol/>
- World Health Organization. (07 de 2019). *eHealth Collaborations*. Obtenido de World Health Organization: <https://www.who.int/ehealth/about/ehealthcollaborations/en/>
- Decreto con Fuerza de Ley N°1. (23 de 09 de 2005). Decreto con Fuerza de Ley N°1 Ministerio de Salud. Diario Oficial de la República de Chile.
- Decreto N°136. (21 de 04 de 2005). Decreto N°136, Reglamento Orgánico del Ministerio de Salud, Ministerio de Salud. *Diario Oficial de la República de Chile*.
- Código Sanitario. (31 de 01 de 1968). Código Sanitario, Ministerio de Salud Pública. *Diario Oficial de la República de Chile*.
- Decreto N°136. (21 de 04 de 2005). Decreto N°136, Reglamento Orgánico del Ministerio de Salud, Ministerio de Salud. *Diario Oficial de la República de Chile*.
- MINSAL. (03 de 2018). *División de Planificación Sanitaria*. Obtenido de MINSAL: <https://www.minsal.cl/wp-content/uploads/2018/03/Presentaci%C3%B3n.-SSP-DIPLAS-Divisi%C3%B3n-de-Planificaci%C3%B3n-Sanitaria.pdf>
- Unidad de Estadísticas y Estándares, DEIS. (2011). *Situación de los Estándares de Información en Salud*. Obtenido de MINSAL: https://repositoriodeis.minsal.cl/deis/estandares/Est%C3%A1ndares_DEIS_2011.pdf
- SSMAULE. (08 de 2014). *La Informática al Servicio de la Salud: El Sistema SISMAULE*. Obtenido de Servicio de Salud Maule: <https://www.ssmaule.cl/minsal/?p=4056>
- Salud-e. (07 de 2015). *Sistemas de Información Servicio de Salud del Maule*. Obtenido de Salud-e: <http://www.salud-e.cl/wp-content/uploads/2015/07/SIS-Maule-Expo-Hospital-2015-R.-B%C3%A4chler-Revisada.pdf>
- Weber, G., Mandl, K., & Kohane, I. (2014). Finding the Missing Link for Big Biomedical Data. *Jama*, 311(24), 2479-2480.

- Mauro, A. (05 de 2019). *Aquí es Donde Viven el 60% de los Datos de Salud de las Personas que no Está en los Sistemas de Información*. Obtenido de Foro Salud Digital: <https://discourse.forosaluddigital.cl/t/aqui-es-donde-viven-el-60-de-los-datos-de-salud-de-las-personas-que-no-esta-en-los-sistemas-de-informacion/389>
- Diana, A. (09 de 2014). *Who Owns EHR Data?* Obtenido de Information Week: <https://www.informationweek.com/healthcare/electronic-health-records/who-owns-ehr-data/d/d-id/1307043>
- SSMC. (07 de 2019). *Proyecto SIDRA*. Obtenido de Servicio Salud Metropolitano Central: <https://www.ssmc.cl/quienes-somos/mision-vision-funciones/proyecto-sidra/>
- Revista Gerencia. (11 de 2012). *Plan Estratégico de TIC del Minsal*. Obtenido de Revista Gerencia: <http://www.emb.cl/gerencia/articulo.mvc?xid=196&sec=7>
- Salud-e. (01 de 03 de 2016). *Servicios de Salud con Desarrollos Propios Construirán Diagnóstico y Mapa de Ruta para Avanzar con SIDRA 2.0*. Obtenido de Ministerio de Salud: <http://www.salud-e.cl/prensa/servicios-de-salud-con-desarrollos-propios-construiran-diagnostico-y-mapa-de-ruta-para-avanzar-con-sidra-2/>
- HL7. (07 de 2019). *Introduction to HL7 Standards*. Obtenido de HL7: <http://www.hl7.org/implement/standards/index.cfm?ref=nav>
- Economía y Negocios. (07 de 2017). *SIDRA II, el Plan Tecnológico que el Minsal Desechó y que, en su Primera Etapa, Generaba Ahorros por \$294 mil Millones*. Obtenido de Economía y Negocios Online: <http://www.economiaynegocios.cl/noticias/noticias.asp?id=379327>
- Salud-e. (04 de 2018). *Ministerio de Salud*. Obtenido de La Estrategia Digital en Salud es un Camino a Largo Plazo que Trasciende a los Gobiernos; Es una Visión de Estado: <http://www.salud-e.cl/prensa/la-estrategia-digital-en-salud-es-un-camino-a-largo-plazo-que-trasciende-a-los-gobiernos-es-una-vision-de-estado/>
- CENS. (12 de 2018). *Ministerio de Salud Avanza en Historia Clínica Compartida Junto al Apoyo de CENS*. Obtenido de Centro Nacional en Sistemas de Información en Salud: <https://cens.cl/minsal-avanza-en-hcc-junto-a-cens/>
- Resolución Exenta N° 347. (21 de 02 de 2019). Resolución Exenta N° 347, División Jurídica. *Ministerio de Salud*. Santiago.
- Hospital Digital. (07 de 2019). *Hospital Digital*. Obtenido de Hospital Digital: <https://www.hospitaldigital.gob.cl/>
- Salud-e. (07 de 2019). *Contexto*. Obtenido de Ministerio de Salud: <http://www.salud-e.cl/plan/contexto/>
- Resolución Exenta N° 595. (02 de 07 de 2019). Resolución Exenta N° 595. Subsecretaría de Redes Asistenciales. *Ministerio de Salud*. Santiago.
- Cluster Salud. (03 de 2019). *Chile: Contraloría Investiga Proyecto "Hospital Digital" del MINSAL*. Obtenido de América Economía Cluster Salud: <https://clustersalud.americaeconomia.com/sector-publico/chile-contraloria-investiga-proyecto-hospital-digital-del-minsal>

- Mauro, A. (11 de 2018). *Registro Unificado de Pacientes (Lista de Espera)*. Obtenido de Foro Salud Digital: <https://discourse.forosaluddigital.cl/t/registro-unificado-de-pacientes-lista-de-espera/34>
- MINSAL. (07 de 2019). *Mi Cirugía*. Obtenido de MINSAL: <https://www.minsal.cl/micirugia/>
- MINSAL. (07 de 2019). *Monitoreo Lista de Espera*. Obtenido de MINSAL: <https://www.minsal.cl/monitoreolistadeespera/>
- El Mercurio. (09 de 2016). Sistema Creado en Chile Predice Riesgo para la Salud de Lactantes Hospitalizados.
- Ríos, S. (07 de 2019). *Almohadita - Ámbar*. Obtenido de Sebastián Ríos Ph. D CV: <http://www.dii.uchile.cl/~srrios/>
- Almohadita. (07 de 2019). *Home*. Obtenido de Almohadita: <http://www.almohadita.cl/>
- Salud-e. (07 de 2019). *Arquitectura Sectorial y Estándares*. Obtenido de MINSAL: <http://www.salud-e.cl/plan/arquitectura-sectorial-y-estandares/>
- Contraloría. (2019). *Informe Final 187-2109 Subsecretaría de Redes Asistenciales Sobre Auditoría a las Adquisiciones de Servicios Informáticos Efectuadas por la Subsecretaría Junio 2019*.
- Date, C. (2000). *Introduction to Database Systems*. Addison-Wesley.
- DEIS. (07 de 2019). *Resúmenes Estadísticos Mensuales DEIS*. Obtenido de Departamento de Estadísticas e Información en Salud: <http://www.deis.cl/resumenes-estadisticos-mensuales-deis/>
- Ballou, D., Wang, R., Pazer, H., & Tayi, G. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, 44(4), 462-484.
- Strong, D., Lee, Y., & Wang, R. (1997). Data Quality in Context. *Communications of the ACM*, 40(5), 103-110.
- Wang, R., & Strong, D. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12(4), 5-33.
- Codd, E. (1989). Relational Database: A Practical Foundation for Productivity. *Readings in Artificial Intelligence and Databases*, 60-68.
- FOLDOC. (12 de 2003). *Free On-Line Dictionary of Computing*. Obtenido de Data Type: <http://foldoc.org/data+type>
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The Real-Time Informational Content of Macroeconomic Data. *Journal of Monetary Economics*, 55(4).
- Stankovic, J., Son, S., & Hansson, J. (1999). Misconceptions About Real-Time Databases. *Computer*, 32(6).
- BioBio Chile. (07 de 2019). *Alertan que Ciberataque Afectó a Empresa Encargada del Servicio de Imágenes en Hospitales*. Obtenido de BioBioChile.cl: <https://www.biobiochile.cl/noticias/nacional/chile/2019/07/10/alertan-que-ciberataque-afecto-a-empresa-encargada-del-servicio-de-imagenes-en-hospitales.shtml>
- Brandom, R. (05 de 2018). *Everything you Need to Know About GDPR*. Obtenido de The Verge: <https://www.theverge.com/2018/3/28/17172548/gdpr-compliance-requirements-privacy-notice>

- Contraloría. (2018). *Informe Consolidado (500-2018) Auditoría al Endeudamiento Hospitalario - Abril 2018*. Santiago.
- Contraloría. (2018). *Informe Consolidado (500-2018) Auditoría al Endeudamiento Hospitalario - Abril 2018*.
- Lalonde, M. (1974). *A New Perspective on the Healths of Canadians*. Ottawa, Canadá.
- BID. (2019). *Comprar, construir o adaptar ¿Cómo decidir?*
- Cortés, M., & Quiroz, E. (2019). *Iniciativas Estructurales para un Crecimiento Costo-Efectivo en el Sistema de Salud*.
- Equipo IBM. (2016). *Arquitectura Empresarial Sector Salud*.
- Wired. (2013). *French National Police Switch 37,000 Desktop PCs to Linux*. Obtenido de <https://www.wired.com/2013/09/gendarmerie-linux/>
- ZDNet. (2008). *ZDNet*. Obtenido de <https://www.zdnet.fr/actualites/la-gendarmerie-passera-ses-70-000-pc-sous-ubuntu-d-ici-a-5-ans-39377943.htm>
- Miranda, S., Teodoro, D., Wei-Kleiner, F., Sundvall, E., Karlsson, D., & Lambrix, P. (2016). Comparing the Performance of NoSQL Approaches for Managing Archetype-Based Electronic Health Record Data. *PLoS one*, 11(3).
- Ramakrishnan, R., & G, J. (2000). Database Management Systems.
- Date, C. (2004). *An Introduction to Database Systems, 8th Edition* (Vol. 8). Pearson.
- Pascal, F. (2012). *dbazine.com*. Obtenido de Nulls Nullified: <http://www.dbazine.com/ofinterest/oi-articles/pascal27/>
- Topol, E. (2019). *Preparing the healthcare workforce to deliver the digital future the Topol Review: An Independent Report on Behalf of the Secretary of State for Health and Social Care*.
- Mauro, A. (04 de 2019). *Por Qué NO Hay que Utilizar la CIE en una Historia Clínica Electrónica*. Obtenido de Foro Salud Digital: <https://discourse.forsaluddigital.cl/t/por-que-no-hay-que-utilizar-la-cie-en-una-historia-clinica-electronica/365?u=alemauro>
- CIPER. (2016). *Grave Falla en la Red del Minsal Dejó Expuesta Información Confidencial de Pacientes*. Obtenido de CIPER: <https://ciperchile.cl/2016/03/05/grave-falla-en-la-red-del-minsal-dejo-expuesta-informacion-confidencial-de-pacientes/>