

# Una mirada a la cultura de datos en Chile



---

**DESAFÍOS Y LECCIONES DE LA  
PANDEMIA POR COVID-19**

DICIEMBRE 2020





## AUTORES

### *Ordenados alfabéticamente*

#### **Demián Arancibia**

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación & Data Observatory

#### **Rafael Araos**

Ministerio de Salud y Facultad de Medicina Clínica Alemana Universidad del Desarrollo, Núcleo Milenio para la Investigación Colaborativa en Resistencia Antimicrobiana (MICROB-R) y Advanced Center for Chronic Diseases (ACCDiS)

#### **Leonardo Basso**

Instituto Sistemas Complejos de Ingeniería & Universidad de Chile

#### **Loreto Bravo**

Instituto de Data Science, Facultad de Ingeniería, Universidad del Desarrollo

#### **José A. Guridi**

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación

#### **Steffen Härtel**

Laboratory for Scientific Image Processing (SCIAN-Lab), Centro de Informática Médica y Telemedicina (CIMT), Centro Nacional en Sistemas de Información en Salud (CENS), Biomedical Neuroscience Institute (BNI), ICBM, Facultad de Medicina, Universidad de Chile

#### **Alejandro Jara**

Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile & ANID - Programa Iniciativa Científica Milenio - Núcleo MiDaS: Centro para el Descubrimiento de Estructuras en Datos Complejos

#### **Alejandro Maass**

Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático & Centro para la Regulación del Genoma, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile - IRL CNRS 2807

#### **Pablo A. Marquet**

Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile. Instituto de Ecología y Biodiversidad (IEB) y Instituto de Sistemas Complejos de Valparaíso (ISCV)

#### **Soraya Mora**

Fundación Ciencia & Vida y Universidad San Sebastián

#### **Tomás Pérez-Acle**

Fundación Ciencia & Vida, Centro Interdisciplinario de Neurociencia de Valparaíso y Universidad San Sebastián

## REVISORES EXTERNOS

#### **Rolando Ocampo**

Director de la División de Estadísticas, CEPAL

#### **Pablo Orefice**

Asesor senior en salud digital, BID  
Director Salud.uy Presidencia. Uruguay

#### **Macarena Rojas**

Pontificia Universidad Católica de Chile

#### **Carlos Sans**

Jefe Departamento de Estadísticas e Información de Salud  
Ministerio de Salud

DICIEMBRE 2020

# CONTENIDOS

---

1	Propósito	4
2	El ciclo de los datos COVID-19	7
	Generación, obtención y acceso	13
	Almacenamiento y curatoría	15
	Análisis, proceso y tratamiento	18
	Visualización	20
3	Brechas, Barreras y Acciones	24
	Brechas y propuestas con foco en el área de salud	26
4	Conclusiones	31

1.

# PROPÓSITO



# PROPÓSITO

---

El valor de los datos ha tomado especial relevancia durante la crisis sanitaria producto de la pandemia por COVID-19. En respuesta a esto, en marzo de 2020 el Ministro de Ciencia, Tecnología, Conocimiento e Innovación, el Dr. Andrés Couve Correa, convocó a diversos actores con el fin de "organizar y disponer datos de la epidemia de COVID-19 para realizar análisis predictivos, científicos y clínicos, contribuyendo su comprensión y a la toma de decisiones basadas en evidencia."<sup>1</sup> Los datos epidemiológicos son recolectados y registrados diariamente gracias al esfuerzo del personal de toda la red de salud y el trabajo de los funcionarios del Ministerio de Salud.

En este contexto, considerando los aspectos administrativos, legales y políticos que la mesa ha enfrentado, se elabora este documento con el objetivo de relevar—a partir de la experiencia de colaboración—la necesidad de un marco apropiado para establecer una gobernanza adecuada y desarrollar capacidades tecnológicas para capturar/producir, procesar y disponibilizar los datos en nuestro país.

El ejercicio de la submesa de datos COVID-19 tiene como propósito el desarrollo de políticas públicas basadas en evidencia. De esta forma, se establece que los descubrimientos y datos son resultados de la colaboración y, por lo tanto, deben estar al servicio de la sociedad, ser transparentes al juicio público y a los desafíos y réplicas que le otorgan validez al quehacer científico.

Desde una perspectiva práctica, la disposición pública de datos debe ser construida en base a estándares internacionales que garanticen su acceso, protección, calidad e interoperabilidad. Esta es la única forma de contar con un ecosistema nacional de datos que favorezca efectivamente la investigación, la innovación, el emprendimiento y el desarrollo económico sostenible. Esto convierte los datos en un activo estratégico y en un acervo cultural que debe ser curado, protegido y cautelado para las futuras generaciones.

---

<sup>1</sup> Sub-mesa de datos COVID-19, Informe 1.

Lo anterior ha poblado las discusiones académicas, económicas, políticas y sociales con conceptos como la “explosión de los datos”<sup>2</sup>, el “diluvio de datos”<sup>3</sup>, el “Big-data”<sup>4</sup>, la “internet de las máquinas”<sup>5</sup>, la “ciber-infraestructura”<sup>6</sup>, la “inteligencia artificial”, el “aprendizaje de máquina”<sup>7</sup>, entre muchos otros. Esto ha fortalecido la relevancia de los datos como un activo estratégico, al ser el insumo básico de un sinnúmero de procesos y tecnologías, pero también al volverse relevantes para los ciudadanos tanto en su uso como en el resguardo de sus derechos. Así, los datos y las tecnologías que emergen con ellos—como la inteligencia artificial moderna—tienen propósitos generales, influyendo múltiples ámbitos de la vida de las personas y habilitando tanto oportunidades como tensiones y dilemas que deben abordarse. Además, el desarrollo tecnológico y los datos no son meros desafíos tecnológicos, sino que, por su ubicuidad en la vida moderna, requiere de instituciones y sociedades preparadas para asumirlos y enfrentarlos como sistemas socio-técnicos<sup>8</sup>.

A continuación se presenta un análisis de la experiencia de la submesa de datos constituida para enfrentar la pandemia por COVID-19, identificando brechas en el ecosistema nacional y proponiendo acciones para contribuir a mejorar la vida de las personas y el desarrollo de nuestro país.

**2** Sweeney, Latanya. “Information explosion.” Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies (2001): 43-74

**3** The Data Deluge. *Nature Cell Biology* 14, 775-775 (2012).

**4** Marx, V. The big challenges of big data. *Nature* 498, 255-260 (2013). <https://doi.org/10.1038/498255a>

**5** IoT overview. International Telecommunications Union. 2012-06-15.

<http://handle.itu.int/11.1002/1000/11559>

**6** Craig A. Stewart, Stephen Simms, Beth Plale, Matthew Link, David Y. Hancock, and Geoffrey C. Fox. 2010. What is cyberinfrastructure. In Proceedings of the 38th annual ACM SIGUCCS fall conference: navigation and discovery (SIGUCCS '10). Association for Computing Machinery, New York, NY, USA, 37-44. DOI:<https://doi.org/10.1145/1878335.1878347>

**7** Russell, Stuart J.; Norvig, Peter (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall. ISBN 978-0-13-604259-4.

**8** Rice, A. K. (1953). Productivity and social organization in an Indian weaving shed: An examination of some aspects of the socio-technical system of an experimental automatic loom shed. *Human Relations*, 6(4), 297-329.

**2.**

## EL CICLO DE LOS DATOS COVID-19

---

# EL CICLO DE LOS DATOS COVID-19

---

La actual pandemia ha relevado la necesidad de contar con una gobernanza adecuada—que permita el acceso, intercambio e integración (All)—de datos para la toma de decisiones basada en evidencia.

Contar con un sistema adecuado de All permite tanto una gestión pública más eficaz a través de la incorporación de nuevas tecnologías, como beneficios en el ámbito científico y económico que pueden generar, en forma oportuna y relevante, conocimiento. Esto aprovecha y potencia eficazmente el conocimiento distribuido en nuestra sociedad siendo capaz de anticiparse ante posibles crisis y desafíos futuros.

El All de datos se potencia por el desarrollo de iniciativas y tecnologías que amplíen el uso de herramientas de análisis de datos, lo que podría repercutir en mejor calidad de vida de las personas y el desarrollo sostenible del país. Un ejemplo es la inteligencia artificial, utilizada en diversos ámbitos como la minería, la agricultura, la salud o la industria alimentaria, promoviendo soluciones innovadoras que nos acercan a metas como los Objetivos de Desarrollo Sostenible de las Naciones Unidas<sup>9</sup>.

Sin embargo, el All de datos en Chile enfrenta múltiples obstáculos y las iniciativas que buscan fomentarlo no se han desarrollado con la velocidad y eficacia necesaria. Individuos y organizaciones—públicas y privadas—se enfrentan a diversos problemas al momento de acceder a datos, incluso al interior de una misma institución y/o sector por falencias técnicas, los marcos jurídicos vigentes—muchas veces llenos de ambigüedades e incertidumbre—e incluso la cultura, tanto de las organizaciones como del país.

Adicionalmente, quienes trabajan con datos se enfrentan a dilemas intrínsecos como la seguridad, la privacidad de los datos o los intereses comerciales. En esta línea, se hace urgente la actualización de la normativa vigente que lleva años detenida—ejemplo de eso es el Proyecto de Ley de Protección y Tratamiento de Datos Personales y la regulación en torno a la ciberseguridad—y el impulso de normas más ágiles que se ajusten a la velocidad que se desarrolla la tecnología, reduzcan la incertidumbre, y generen una gobernanza de datos coherente entre los sectores económicos, el gobierno, la sociedad, y el contexto internacional.

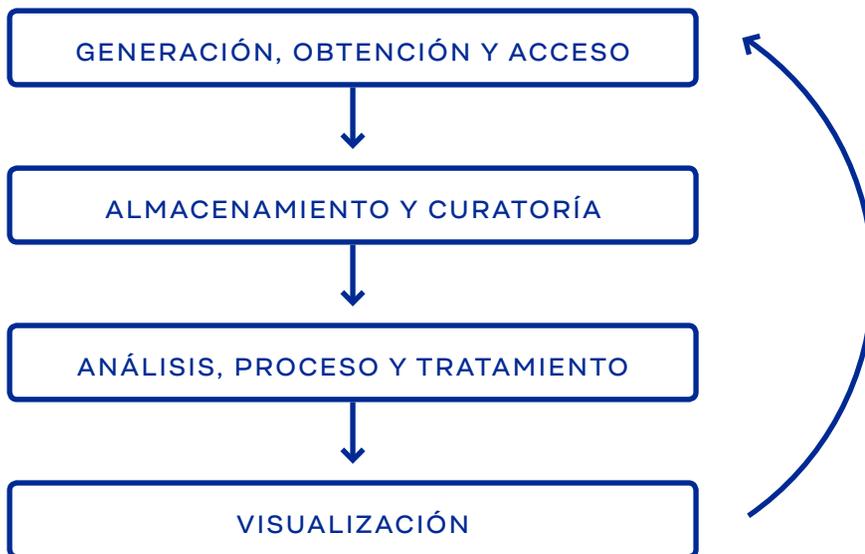
---

<sup>9</sup> Ver en <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>

Esto es especialmente importante en momentos de crisis, en los que se requieren dinámicas colaborativas que den paso a soluciones rápidas basadas en datos, y no transformen su uso en conflictos que no contribuyen a la resolución de problemas.

Para poder profundizar en las brechas y acciones que, a partir de la experiencia de la pandemia por COVID-19, identificamos y proponemos como grupo, es conveniente analizar los beneficios y riesgos asociados al All de datos durante todo su ciclo de vida (Figura 1).

FIGURA 1  
El Ciclo de los Datos. Elaboración propia



Por tanto, para combatir la pandemia por COVID-19 de forma eficaz se requiere de datos que puedan proporcionar información sobre las características, evolución, prevalencia, impactos y sus consecuencias. En particular, se han requerido datos de múltiples fuentes y características (ej. epidemiológicos, movilidad, medio ambiente). En esta línea, no basta con la obtención de datos, sino que estos deben cumplir una serie de características como accesibilidad, actualidad y completitud (ver Figura 2) que se ajusten al tipo de dato con el que se trabajará. Para los fines de esta publicación, proponemos cuatro clasificaciones para categorizar los datos y que, dependiendo de esta, puede requerir distintos estándares de calidad:

- **Según tópico:** Se refiere a la temática que abordan. Ejemplos: Transporte, Salud, Seguridad, Agricultura, Educación, Energía, etc.
- **Según formato:** Se refiere a la forma que tienen los datos. Estos pueden ser estructurados o no estructurados. Ejemplos: Audio, Video, Imágenes, Textos, CSV, etc.
- **Según origen:** Se refiere a la fuente de los datos. Ejemplos: Servicios Públicos, Empresas Privadas, Grupos Académicos, etc.
- **Según tipo de acceso:** Se refiere a quién puede acceder a los datos y de qué forma. Ejemplo: Datos Abiertos, Restringidos a académicos, etc.

FIGURA 2  
15 Dimensiones en Calidad de Datos.



**Fuente:** Araya, Laura (2020). "Modelo para la optimización de la captura y validación de datos para la fiscalización de las oportunidades de la Garantía Explícita en Salud (GES) en Isapres".

Así, la necesidad de disponer datos no se reduce a su disponibilidad, sino que a generar procesos, protocolos y buenas prácticas que faciliten el manejo de estos por múltiples usuarios.

En el caso de la pandemia por COVID-19, dado que el principal vector del virus es el ser humano y que la pandemia se ha desarrollado en forma distinta a epidemias anteriores, el requerimiento de datos adecuados es indispensable para responder a la emergencia con eficacia en un lapso de tiempo acotado. Para lograrlo, fue necesario generar políticas y procesos de apertura de datos en todo el mundo con distintos niveles de acceso, variedad, cantidad y calidad de datos. Esto ha permitido generar análisis y modelos matemáticos que permiten caracterizar la propagación del virus y apoyar la toma de decisiones. El nivel de eficacia de los modelos y recomendaciones ha estado sujeto a que las bases de datos estén debidamente curadas, que se actualicen con frecuencia y en qué medida los expertos pueden lidiar con datos parciales o de mala calidad, entre otros.

En el caso de Chile, los datos epidemiológicos y de movilidad han resultado fundamentales para el análisis de la propagación y severidad de la infección por COVID-19. Los datos utilizados incluyen, entre otros:

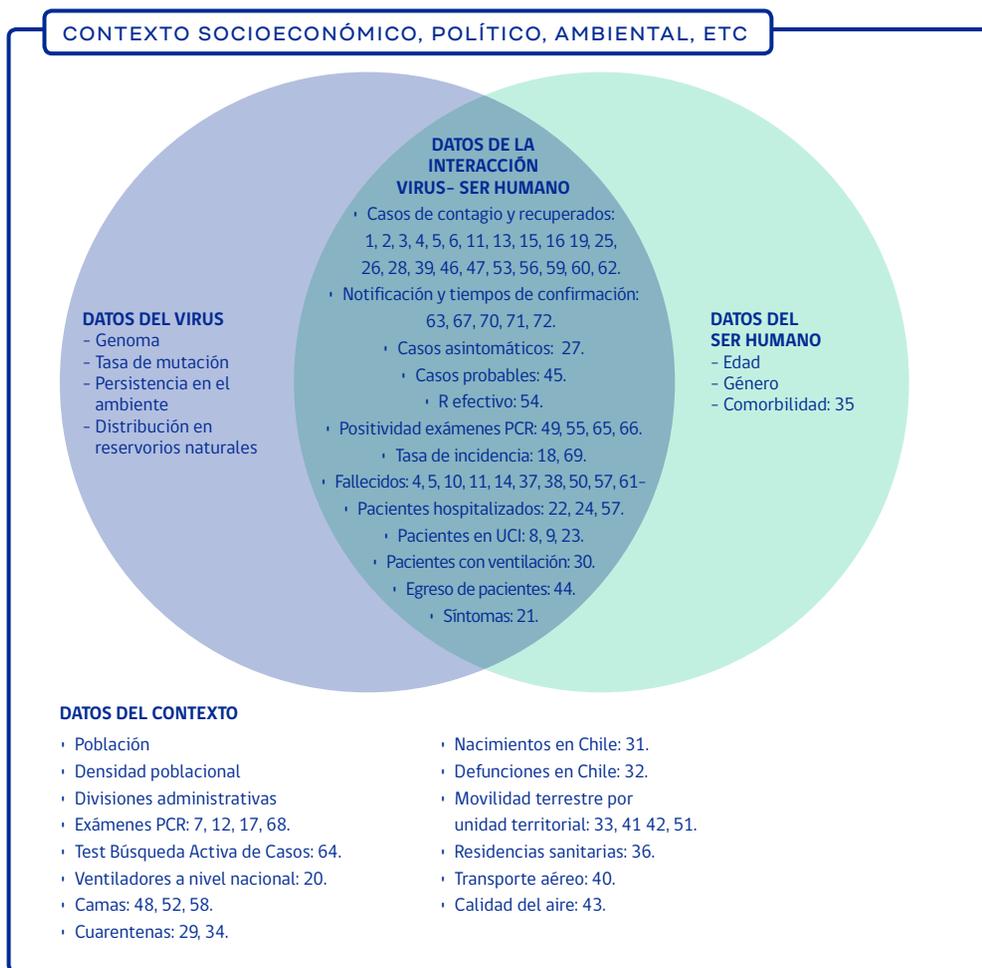
1. Sexo, comorbilidad, y rango etario de contagiados y fallecidos.
2. Cantidad de contagios nuevos por día, síntomas, tasa de positividad y otros parámetros de evolución clínica de los casos.
3. Movilidad de las personas, áreas de propagación, etc.

Estos datos han permitido múltiples análisis que apoyan en la toma de decisiones sanitarias como las cuarentenas en distintos sectores, el distanciamiento social, el uso de mascarilla en la vía pública, el testeo dirigido, entre otras. Además, son cruciales para enfrentar la etapa de desconfinamiento en forma eficaz. Este análisis es posible en la medida que existe acceso a los datos, tanto por quienes cuentan con las capacidades técnicas para desarrollar los modelos y su análisis, como por la población en general en pos de transparencia y *accountability*. Esto es un círculo dinámico que aprende en su desarrollo y que debe ser gestionado adecuadamente.

Agrupando los datos según su fuente, tenemos datos del contexto en el que se desenvuelven las personas, datos propios del agente patógeno, datos propios de las personas y datos que resultan de las interacciones entre estos. En forma resumida, la figura 3 presenta los datos que se han disponibilizado a la fecha de publicación de este documento según su proveniencia.

**FIGURA 3**  
**Productos de datos disponibles a la fecha de publicación según su fuente, cuyo total asciende a 72. Se simplificó el detalle de los datos, agrupando en categorías, pero cada producto representa distintas aproximaciones y procedimientos.**

Elaboración propia.



No obstante, para lograr pasar de la fuente de datos al apoyo de la toma de decisiones, existe una dinámica de transformación que describiremos a continuación, en lo que llamamos anteriormente el Ciclo de los Datos (figura 1).

# 1. Generación, obtención y acceso

La primera etapa del ciclo consiste en la generación u obtención y acceso a los datos. Este proceso está limitado por factores tecnológicos (ej. sensores), humanos (ej. competencias, motivación), organizacionales (ej. procesos) e institucionales (ej. normativa). El estado de estos factores puede incluso forzar a que se combinen etapas del ciclo antes que los datos sean accesibles, produciendo, en ocasiones, lagunas temporales entre la generación/obtención y el acceso. Así, la dinámica que vivirá cada usuario será diferente. Un ejemplo en el contexto de la pandemia fueron los datos de movilidad, que pertenecían a organismos privados y debieron ser procesados (etapa de análisis) para poder disponibilizar una forma de ellos.

Una característica particular que tienen los datos utilizados durante la pandemia es que su generación está fuertemente distribuida. La información puede provenir de los pacientes (epidemiológicos), de su comportamiento de movilidad, de su comportamiento de consumo energético, entre otros. Esto hace que existan múltiples fuentes y propietarios de los datos donde cada uno tiene sus respectivos formatos, protocolos y estándares. Así, el acceso en forma holística requiere coordinación y controles de calidad, en especial cuando los datos fluyen entre organizaciones.

Por otro lado, al momento de obtener/generar los datos, se debe tener en cuenta la regulación asociada. Esto es especialmente importante cuando se trata de datos de personas, como es el caso de la pandemia, ya que estos deben ser cautelados por su carácter sensible y privado. Así, encontrar el equilibrio entre la protección de los datos personales, el acceso a ellos por parte del Estado para la toma de decisiones y la generación de información pública para generar confianza en la población, es sumamente delicado.

La actual pandemia nos enfrentó a una serie de dificultades técnicas y regulatorias en la obtención, generación y acceso a los datos en forma simultánea. Sin embargo, muchas de estas dificultades pudieron resolverse exitosamente gracias a el esfuerzo de los actores de la red de salud, la colaboración entre múltiples actores públicos y privados y la incorporación y actualización de herramientas tecnológicas.

Para lograr contar con estos datos, el desafío de la accesibilidad toma especial relevancia. De hecho, en el caso de Chile, muchos datos existían (ej. datos de las empresas de telecomunicación), y eran utilizados en ámbitos ajenos al interés del Estado, sin embargo, no eran accesibles para las autoridades y/o los científicos. La razón esgrimida para esta falta de acceso usualmente tenía relación con la protección de datos personales. De hecho, en el caso del Ministerio de Salud (Minsal), solo dos departamentos tienen acceso a datos sensibles (de Epidemiología y de Estadísticas), con fuertes restricciones incluso para el intercambio al interior del Estado.

En este punto existe una curiosa paradoja: los datos epidemiológicos utilizados por los tomadores de decisiones son fuertemente exigidos con gran desagregación y en el estado más crudo posible—pese a ser personales—, argumentando la necesidad de transparencia. No obstante, aquellos en manos de privados, solo se entregaron a la autoridad mediante convenios puntuales y en un estado procesado, lo que se acepta con naturalidad por parte de la ciudadanía. Así, existen diferencias con respecto a los derechos y el uso de datos cuando son manejados por el sector público y el privado, las que muchas veces son poco claras, forzando la formulación de soluciones de acceso *ad-hoc*. En el caso de la información epidemiológica recopilada y manejada por el Ministerio de Salud, el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (Minciencia) jugó un rol crucial para su acceso por parte de la academia vía convenios de colaboración. En el caso de los datos de empresas privadas fue vía convenios entre privados.

El resultado final, si bien adecuado, logró disponer datos valiosos para la toma de decisiones, el monitoreo de la emergencia y el incremento de la seguridad y confianza en las cifras, se hizo en forma apurada, sin protocolos en qué basarse y dependiendo de voluntades individuales. Un sistema virtuoso de flujo de AII de datos en emergencia debe contar con una institucionalidad y con protocolos de acceso y disposición que no den pie a arbitrariedades en la interpretación en momentos de crisis. Además, se requieren mecanismos ágiles para la colaboración entre el Estado, la academia, la industria y la sociedad civil, que les permitan responder rápida y eficazmente a las crisis. Solo así es posible contar con datos de múltiples fuentes públicas y privadas que interoperan y contribuyan al bienestar de las personas en una emergencia.

## 2. Almacenamiento y curatoría

La segunda etapa del ciclo tiene relación con la curatoría y el almacenamiento de los datos, procesos indispensables para la disposición y generación de valor. De esta forma, los datos se procesan para cumplir con estándares de calidad e interoperabilidad reconocidos. Esto puede ocurrir en forma manual, semi-automatizada y automatizada, siendo lo último el ideal para minimizar los errores humanos y la actualidad de los datos. No obstante, se requieren sistemas que permitan a los usuarios reportar errores, vigilando la calidad de los datos y el funcionamiento de los procesos.

Para un adecuado almacenamiento y curatoría, se requiere contar con un repositorio que permita disponer e integrar datos, e interactuar con la comunidad. En esta línea, el Minciencia tomó la decisión junto al Data Observatory de disponer los datos en la plataforma GitHub como repositorio central.

### DATA OBSERVATORY (DO)

Iniciativa público-privada sin fines de lucro cuyo fin es potenciar el beneficio que obtenemos de los datos públicos, de valor global y únicos que se están generando en nuestro país. Articula a la academia, sector público, industria y sociedad civil en alianzas globales para generar soluciones y capacidades en ciencia de datos y tecnologías afines que sean útiles y tengan retornos en diversos sectores de la economía.

El DO se centra en cuatro líneas de trabajo para cumplir su misión: (i) reunir y poner a disposición del público conjuntos de datos de valor global de manera abierta; (ii) diseñar e implementar soluciones para adquirir, analizar, explorar, visualizar y ofrecer acceso a estos conjuntos de datos y acelerar su aprovechamiento; (iii) contribuir a la formación de talento relacionado con la ejecución de estas acciones, y articular material y cursos a partir de experiencias de primera mano; y (iv) invertir en la creación de redes para facilitar transferencia tecnológica y asociatividad entre personas que trabajan en campos distintos pero que tienen una relación de similitud funcional en sus áreas de trabajo con datos.

Durante la Pandemia por COVID-19, el DO ha contribuido activamente apoyando técnicamente la disposición de datos en el repositorio abierto del Minciencia.

Ver más en [www.dataobservatory.net/](http://www.dataobservatory.net/)

### GITHUB

Es una plataforma de desarrollo de software colaborativa creado en 2008, en el que las personas pueden compartir su código y trabajar en conjunto con un sistema de control de versiones. Actualmente es la plataforma más importante de colaboración en proyectos Open Source.

Pese a haber sido originalmente diseñado para compartir código, ha sido destacado por la revista Nature como una herramienta que permite democratizar el acceso a bases de datos<sup>10</sup> por sus características de manejo abierto, transparente, con control de cambios (que registra cualquier cambio hecho a datos y código) y que permite la interacción con la comunidad.

Ver GitHub creado por el Minciencia en: [github.com/MinCiencia/Datos-COVID19/](https://github.com/MinCiencia/Datos-COVID19/)

---

<sup>10</sup> Perkel, J. (2016). Democratic databases: science on GitHub. Nature News, 538(7623), 127.

Durante el trabajo de la submesa de datos, se pasó de procesos principalmente manuales de actualización a automatizar la actualización de la mayoría de los data products disponibles en el repositorio mediante herramientas como Python y Github Actions. Para esto, en abril de 2020 se construyó un producto mínimo viable que consistía en la elaboración de archivos CSV (archivo separado por comas) a partir de los informes de epidemiología del Minsal en formato PDF. El piloto fue sumamente exitoso, generando miles de clones del repositorio (copias completas) y visitantes a los distintos productos de datos que se fueron produciendo a partir de eso.

Con el paso de los meses, se fueron incorporando nuevos productos de datos, con mayor nivel de desagregación—siempre respetando la privacidad de las personas—, y proveniente de distintas fuentes como el Ministerio de Transporte y Telecomunicaciones (ej., datos de transacciones de la tarjeta de transporte público (BIP!) y de movilidad entre comunas), del Ministerio de Medio Ambiente (ej., datos de calidad del aire), y de empresas a través de grupos académicos (ej., datos de movilidad a través de los datos de telefonía).

Los estándares producidos para los datos epidemiológicos incluyeron la disposición en tres esquemas: basado en filas (cada fila contiene el número de contagiados para todas las fechas), en columnas (cada columna contiene el número de contagiados para todas las fechas), y estandarizados (cada fila da el número de contagiados para una comuna y fecha específica). Además, se verificó que las bases fueran curadas para ser interoperables con otros tipos de datos como de medio ambiente y transporte.

Un aspecto muy importante para la gestión del repositorio y la curatoría de los datos, fue la posibilidad de interactuar con la comunidad a través de dos herramientas dadas por el sistema: pull requests and issues. Los pull requests permiten que cualquier usuario pueda proponer actualizaciones a los datos y los issues que los usuarios informen sobre problemas en los datos tales como duplicaciones, errores de imputación, falta de actualización u otros.

Todas las características antes descritas resultan vitales para tener un sistema de almacenamiento de datos que permita democratizar el acceso a los datos: transparencia, control de versión e interacción con la comunidad.

### 3. Análisis, proceso y tratamiento

Ya sea para la pandemia, otras emergencias, o en general el trabajo con datos, usualmente se utilizan modelos y análisis matemáticos para comprender los procesos y apoyar la toma de decisiones. En el caso de la actual situación sanitaria, se volvió sumamente relevante que la comunidad experta generara modelos con el fin de estimar o predecir el comportamiento del virus en pos anticipar medidas que permitan mitigar su efecto.

Por otro lado, fue necesario crear series de datos procesados sobre los crudos dispuestos por el Minsal. Esto se hizo para hacerse cargo de problemas inherentes a la toma de datos, como son los retrasos en los reportes de qPCRs de las fechas reales de inicio de síntomas u otros efectos de la cadena de producción de datos. Estas series, así como aquellas producidas por los modelos matemáticos constituyen un corpus importante de datos sintéticos que permitieron dar más profundidad a la toma de decisiones.

#### MODELOS MATEMÁTICOS Y COVID-19

Durante el desarrollo de la pandemia por COVID-19 se han utilizado diversos modelos que van desde clásicos llamados de compartimentos (derivados de los modelos de tipo SIER desarrollados por Kermarck y Mckendrick en 1927) hasta micro-locales basados en agentes. En estos modelos, se clasifica a los individuos como susceptibles, infectados, hospitalizados, etc., los cuales interactúan a través de reglas dinámicas que pueden ser deterministas o probabilísticas.

Los distintos modelos han permitido evaluar el comportamiento del virus en el mundo y diseñar indicadores de seguimiento de la pandemia e implementar medidas de mitigación y control. Adicionalmente, gracias a la tecnología actual, se han explorado nuevos modelos basados en ella. Hoy es posible el rastreo de teléfonos móviles y analizar patrones con inteligencia artificial, lo que puede llevar incluso a predecir zonas de riesgo y probabilidades de contagio con mayor eficacia.<sup>11</sup>

**11** Godio, A., Pace, F., & Vergnano, A. (2020). SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence. *International Journal of Environmental Research and Public Health*, 17(10), 3535.

En Chile, se han implementado distintas medidas de mitigación, como son el cierre de colegios, comercio, prohibición de eventos masivos y cuarentenas. De acuerdo al Minsal, los indicadores epidemiológicos que se consideran para la toma de decisiones son (i) el número de infectados activos—individuos que se encuentran contagiados y pueden contagiar—, (ii) la incidencia de casos activos por 100.000 habitantes, (iii) la concentración de infectados activos por km<sup>2</sup> y (iv) la capacidad asistencial de los servicios de salud locales. Estos indicadores se han incorporado en la creación de los distintos modelos, como se puede ver en <https://github.com/Instituto-Milenio-de-Datos/modelamiento-covid>.

Los modelos han podido contribuir a cuantificar la expansión de la pandemia y a la toma de decisiones en la medida que ha existido una generación y obtención confiable de los datos, un tratamiento adecuado de estos, un constante análisis crítico de la metodología y la información recopilada, y la incorporación de variables que consideren el contexto social, económico y geográfico particular del país. Para lograr lo último, se necesitó la integración de datos socioeconómicos, demográficos, entre otros, de distintas instituciones del Estado como el Instituto Nacional de Estadística. Sin embargo, dado el poco tiempo transcurrido y la gran novedad que esta pandemia ha revelado, todavía es necesario un largo trabajo de exploración de los datos generados para eventualmente explicar los fenómenos inherentes a ellos.

## 4. Visualización

Por último, la visualización de los datos, con distintos grados de procesamiento, es fundamental para generar un adecuado flujo de información a la sociedad, orientando la toma de decisiones de mejor manera, y contribuyendo a una mayor transparencia y educación de la ciudadanía. Esto es indispensable en una sociedad democrática, pero es en sí mismo un proceso sumamente complejo.

La pandemia trajo consigo un explosivo desarrollo de distintos tipos de plataformas y herramientas de visualización de la dinámica espacial y temporal del virus. A nivel mundial, la plataforma más utilizada ha sido el John Hopkins Coronavirus Resource Center (<https://coronavirus.jhu.edu/map.html>), sin embargo también existen diversos esfuerzos locales provenientes tanto de la Academia y el mundo privado, como del Estado.

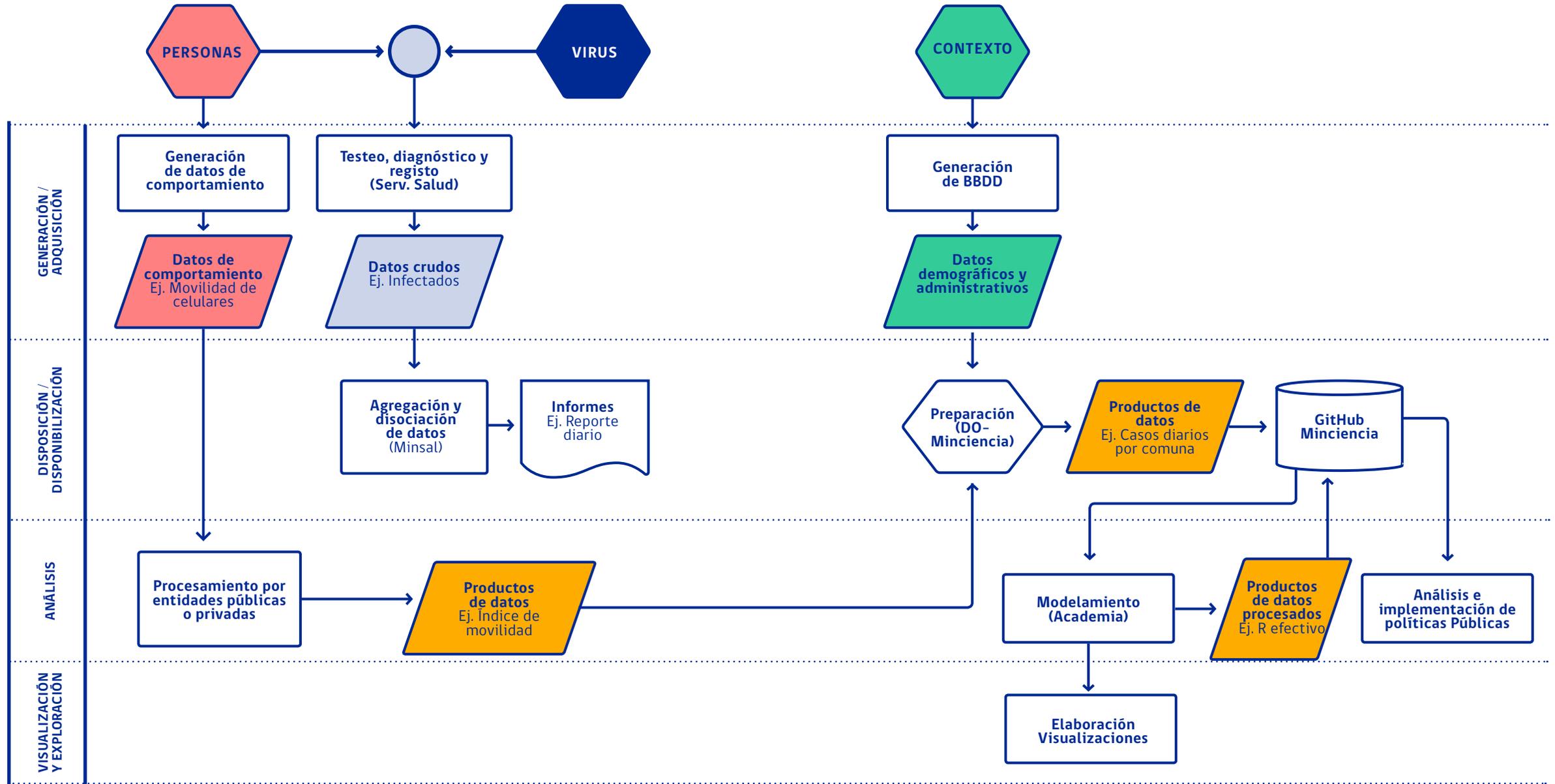
En este punto, es relevante que exista una coordinación entre los actores involucrados en las distintas etapas del ciclo de datos. En el caso de la pandemia, el Minciencia jugó un rol relevante al proveer la plataforma con los datos y APIs necesarias para montar visualizadores, lo que se realizó en coordinación con las fuentes de los datos que, a su vez, establecen condiciones sobre cómo y cuándo se pueden disponer para visualizar. Algunos ejemplos de las desarrolladas en Chile se presentan en la Tabla 1.

Finalmente, en la figura 4 se presentan ejemplos de los flujos de distintos datos a través de ellas, ilustrando la complejidad del ciclo de datos, la multiplicidad de actores existentes y los distintos tipos de procesos involucrados.

**TABLA 1**  
**Ejemplos de visualizadores de datos COVID-19 en Chile.**

NOMBRE DEL VISUALIZADOR	DESCRIPCIÓN	DIRECCIÓN WEB
Plataforma Territorial Ministerio de Salud	Interfaz de visualización de datos de casos COVID-19 confirmados y probables, representados en cuadrantes urbanos de 1 kilómetro cuadrado para todo el país obtenidos de la plataforma EPIVIGILA.	<a href="https://storymaps.arcgis.com/stories/fbf237db339b4422b-51cfbd707db2810">https://storymaps.arcgis.com/stories/fbf237db339b4422b-51cfbd707db2810</a>
Visualizador Covid-19 Chile	Esta aplicación presenta de forma gráfica la información relativa a la evolución de la pandemia del Coronavirus COVID-19 en Chile, a partir del repositorio de la Mesa de Datos.	Esta aplicación presenta de forma gráfica la información relativa a la evolución de la pandemia del Coronavirus COVID-19 en Chile, a partir del repositorio de la Mesa de Datos. <a href="https://coronavirus.mat.uc.cl/">https://coronavirus.mat.uc.cl/</a>
Cifras Oficiales COVID-19	Reporta las cifras oficiales reportadas por el Ministerio de Salud en sus reportes diarios e informe epidemiológico Covid-19.	<a href="https://www.gob.cl/coronavirus/cifrasoficiales/#reportes">https://www.gob.cl/coronavirus/cifrasoficiales/#reportes</a>
Visor Territorial COVID-19 Ministerio de Bienes Nacionales (IDE Chile)	Visualizador de datos territoriales a nivel regional y comunal, casos activos, nuevos, fallecidos etc y estado de cuarentena de las comunas.	<a href="https://www.visorterritorial.cl/">https://www.visorterritorial.cl/</a>
Visualizador COVID-19 (Centro de Modelamiento Matemático, CMM)	Visualizador de datos epidemiológicos básicos, índices, evolución de la pandemia en Chile, comparación a nivel regional y proyección en base a modelos.	<a href="https://covid-19vis.cmm.uchile.cl/geo">https://covid-19vis.cmm.uchile.cl/geo</a>
CovidAnalytics (Instituto Sistemas Complejos de Ingeniería)	Visualizador de movilidad y Visualizador de datos de consumo eléctrico a nivel comunal.	<a href="https://covidanalytics.isci.cl/">https://covidanalytics.isci.cl/</a>
iCOVID CHILE (Plataforma Interuniversitaria, UC, UCH y UCON)	Visualizador de indicadores de Dinámica de contagio, testeo, Capacidad hospitalaria y trazabilidad y aislamiento.	<a href="https://www.icovidchile.cl/">https://www.icovidchile.cl/</a>
COVID-19 Information System for the Region of the Americas (Pan American Health Organization PAHO)	Visualizador de datos epidemiológicos comparativos para las Américas.	<a href="https://paho-covid19-response-who.hub.arcgis.com/">https://paho-covid19-response-who.hub.arcgis.com/</a>

FIGURA 4  
Flujo de datos a través de las cuatro etapas del ciclo para algunos tipos de datos durante la pandemia en Chile. Elaboración propia.



**3.**

BRECHAS,  
BARRERAS Y  
ACCIONES

---

# BRECHAS, BARRERAS Y ACCIONES

---

A partir de la experiencia y el análisis del All de datos, en la siguiente sección se plantean una serie de brechas y barreras que se han identificado, con sugerencias de acciones a tomar desde el Estado, la academia y la sociedad civil. Este análisis no pretende ser exhaustivo, sino que se basa en el trabajo realizado por la submesa de datos en el contexto de la pandemia por COVID-19.

Chile ha tenido una larga tradición estadística. La primera oficina estadística fue formalmente establecida entre los años 1843 y 1851, y su instalación tuvo como objeto generar una institucionalidad que entregara regularidad y método a la recolección y procesamiento de los datos necesarios para conocer y medir los recursos del país. En la actualidad contamos con un sistema estadístico descentralizado, donde el Estado está facultado para usar los datos con fines prestacionales. El artículo 7 de la Constitución de la República vigente establece el marco de acción, indicando quiénes son las autoridades competentes para recolectar y procesar datos. El Instituto Nacional de Estadística (INE) centraliza un número importante de datos, sin embargo, son varios los organismos públicos los que también tienen esta competencia. El Banco Central de Chile lleva a cabo las cuentas nacionales, el Ministerio de Salud las estadísticas sanitarias, el Ministerio de Medio Ambiente las estadísticas ambientales, y el Ministerio de Desarrollo Social datos sobre la caracterización socioeconómica de la población, por mencionar algunos ejemplos.

El sistema actual exhibe una serie de debilidades que se han visto estresadas con la pandemia. Si bien el sistema vigente puede tener un funcionamiento adecuado en una situación de baja carga, la pandemia cambió el tipo y volumen de los datos requeridos, y la velocidad con la que se requiere de disponer evidencia basada en datos. En efecto, el diseño de medidas ajustadas a la realidad nacional ha requerido, y seguirá requiriendo, del análisis de diferentes fuentes de información, y la cuantificación del riesgo actual para la población exige de capacidades específicas en materia de recolección, tratamiento y modelamiento de datos, que, si bien pueden estar disponibles en ciertas reparticiones del estado, ellas no actúan de forma articulada.

La falta de modernización del sistema nacional relacionado a los datos antes de la pandemia tiene diferentes orígenes. Las principales razones son (i) la falta de reconocimiento y comprensión de su necesidad y valor, (ii) la necesidad de una mayor coordinación y presupuesto, y (iii) la poca claridad sobre su definición, objetivos y posibles diseños.

En esta línea, en el quéhacer de la submesa de datos y la respuesta a la pandemia, ha surgido una comprensión del valor de datos para crear evidencia. Así, la pandemia ha tenido un efecto de sensibilización en la sociedad, que puede subsanar los problemas enunciados anteriormente generando una mayor coordinación política que permita realizar modificaciones normativas y presupuestarias para establecer un diseño general eficiente. De hecho, el propio Minciencia ha tenido dificultades para disponibilizar datos, lo que sugiere algunas características que debiese tener el sistema como autonomía.

## BRECHAS Y PROPUESTAS CON FOCO EN EL ÁREA DE SALUD

En 2005, la Asamblea Mundial de la Salud en su resolución WHA58.28 sobre salud digital, insistió a sus Estados Miembros "considerar la posibilidad de elaborar un plan estratégico a largo plazo para desarrollar y poner en práctica servicios de salud digital (...), desarrollar la infraestructura de las tecnologías de la información y las comunicaciones para la salud (...), y promover el acceso equitativo, asequible y universal a sus beneficios."<sup>12</sup> En 2013, la Asamblea aprobó la resolución WHA66.24 sobre normalización e interoperabilidad de la salud digital, con la sugerencia de "considerar la posibilidad de elaborar (...) políticas y mecanismos legislativos vinculados a una estrategia nacional general de salud digital"<sup>13</sup>.

En los últimos 15 años, Chile ha seguido diferentes agendas digitales que tocan la materia, entre ellas la agenda del Libro Azul (2006), la estrategia en Sistemas de Información De la Red Asistencial SIDRA (2008), el marco frustrado de soluciones informáticas SIDRA II (2016), la Plataforma Convergente (2017/18), el Hospital Digital (2018/19), y una serie de acciones orientadas a la contingencia de la crisis sanitaria. El frecuente cambio de estrategias y agendas motivados por los diferentes gobiernos no logra responder a diagnósticos de expertos y no recoge la experiencia internacional que enfrenta el desarrollo en salud digital con estrategias y agendas con horizontes de 5 a 15 años. En esta línea, se ve la necesidad de generar agendas que no dependen del gobierno de turno, sino que sean transversales, independientes, con recursos propios y competencias claras.

Se han diagnosticado múltiples brechas y barreras que aluden a la adopción de tecnologías digitales en base a evidencia en Salud, la región y a nivel global. Estos diagnósticos se han evidenciado dramáticamente en el contexto de la pandemia COVID-19. La realidad nacional en el ámbito de Salud presenta brechas y barreras en las áreas de Capital Humano, Interoperabilidad y Calidad de Datos que requieren acciones, pues constituyen factores habilitantes para la implementación de sistemas de información en salud que contribuyen a la disposición de evidencia basada en datos.

<sup>12</sup> Ver <https://www.who.int/healthacademy/media/WHA58-28-en.pdf>

<sup>13</sup> Ver [https://www.who.int/ehealth/events/wha66\\_r24-en.pdf?ua=1](https://www.who.int/ehealth/events/wha66_r24-en.pdf?ua=1)

## • BRECHAS EN CAPITAL HUMANO

En Chile, la brecha de fuerza laboral en salud digital no está cuantificada en detalle. Estimaciones recientes basadas en el Modelo de Adopción de Registro Clínico Electrónico (EMRAM) plantean tres escenarios a corto, mediano y largo plazo: (i) se requiere una fuerza laboral de 2.320 personas en jornada laboral completa para una implementación adecuada y segura de Sistemas de Información en Salud (SIS)—suponiendo que un 53% del sistema de salud tiene un nivel de adopción mínimo—, (ii) a mediano plazo, se requieren a 8.038 personas capacitadas en Salud Digital—suponiendo que Chile logre una distribución similar de niveles de adopción a la de EEUU 2014—y (iii) a largo plazo se requieren a 15.675 personas capacitadas en el área—suponiendo un escenario ideal con un nivel de adopción óptimo—.

Hoy en día, Chile cuenta con una fuerza laboral escasa y poco reconocida por el medio. En 2019, se elaboró y publicó el primer Modelo Referencial de Competencias en Sistemas de Información en Salud (<https://cens.cl/modelo>) y un Mapa Funcional de Procesos del sector que permitió identificar 18 Perfiles Laborales necesarios, de los cuales 5 fueron levantados y publicados recientemente (<https://cens.cl/perfiles-laborales>). Hasta la fecha, el país no cuenta con un marco de cualificaciones que establezca rutas formativo-laborales que faciliten y permitan la formación de capital humano para el área. En particular, en el caso concreto de la pandemia se hace cada vez más relevante contar con especialistas que reflexionan desde la llamada Epidemiología Digital, es decir desde la tecnología, los datos que ellas generan y los métodos para su análisis<sup>14</sup>.

Más allá, el trabajo con datos y las tecnologías asociadas es profundamente multidisciplinario. No solo se requieren perfiles relacionados a la salud, sino que se necesitan expertos en áreas de ciencia y tecnología, de ciencias sociales, de las humanidades, del derecho, entre otras. Esto es fundamental para aproximarse a los problemas en forma comprensiva, comprendiendo los impactos que existen al trabajar con datos y utilizándolos en forma adecuada más allá de las dimensiones técnicas. El análisis de datos no es simplemente acceder a ellos y modelar, sino que es una tarea compleja, con múltiples técnicas posibles y variadas interpretaciones que deben ser contrastadas.

### ACCIONES PROPUESTAS:

1. Implementar programas que permitan desarrollar competencias a nivel técnico, profesional y de postgrado. Esto debe ser no solo en el campo de datos y tecnología, sino también en ciencias sociales y humanidades relacionadas a estos.
2. Promover la oferta de formación en cursos de capacitación para estimular la reconversión laboral y/o la formación continua de la fuerza laboral.

<sup>14</sup> Salathé M. Digital epidemiology: what is it, and where is it going?. Life Sci Soc Policy. 2018;14(1):1. Published 2018 Jan 4. doi:10.1186/s40504-017-0065-7

3. Desarrollar certificaciones y reconocimiento de las competencias basadas en los Perfiles Laborales elaborados por expertos del área. De esta manera, se contribuye a tener puestos de trabajo para personas calificadas que avanzan con decisión en mejorar la salud del país a través de los datos.

## • BRECHAS EN INTEROPERABILIDAD, ESTÁNDARES Y TERMINOLOGÍAS

En Chile, si bien la implementación de soluciones informáticas en salud ha sido impulsada por las políticas de digitalización del Estado, los sistemas de información en salud no conversan entre ellos, transformando la red de salud nacional en un conjunto de islas de información<sup>32</sup>. Una de las causas fundamentales es la falta de implementación y adopción de un modelo de madurez de salud digital en el ecosistema que permita un desarrollo armónico con mecanismos claros de encauzamiento y medición de capacidades existentes.

Las dificultades en la gestión de información crítica para abordar la pandemia COVID-19 han puesto en evidencia la necesidad de mejora de comunicación en el sistema de salud. Se identifican las carencias de: (i) contar con plataformas tecnológicas robustas y accesibles en todo el territorio, (ii) definiciones y adopción de estándares y terminologías, y (iii) contar con un marco normativo que responda a la necesidad de un sistema de salud público privado. Esto ha impedido una red de salud conectada con empoderamiento (*empowerment*) sobre los datos e información a nivel de salud pública, deteriorando la capacidad efectiva en la toma de decisiones oportunas.

Existen en la actualidad signos en el país que demuestran que el estado de madurez del medio se está desarrollando en torno a potenciar los pilares de interoperabilidad y los facilitadores para poder desarrollar la interoperabilidad en salud: (i) la creación de organizaciones como CENS y HL7 Chile; (ii) desarrollos propios tanto de privados como en Servicios de Salud y Prestadores públicos en relación a terminologías e identificadores de pacientes; y (iii) el desarrollo de proyectos usando estándares de interoperabilidad internacionales por parte del estado. Hoy existe la oportunidad de aprovechar la experiencia internacional en el uso de los estándares y la capacidad de los nuevos estándares de interoperabilidad como HL7 FHIR® que están siendo adoptados a nivel global.

Considerando la importancia de formar y certificar capacidades en nuestro sistema de salud, así como validar la conformidad de desarrollos locales contra buenas prácticas y estándares internacionales, desde el 2018 CENS ha implementado un conjunto de talleres a nivel nacional, enfocado en el aprendizaje práctico entre las diferentes disciplinas que requieren la transformación digital de la salud. Así también un conjunto de proyectos informáticos que permitan diseñar, desarrollar y promover los estándares de interoperabilidad en instituciones públicas y privadas, como cuenta médica interoperable, desarrollo de la primera plataforma latinoamericana para enseñar HL7 FHIR®, o el proyecto Esperanza, que automatiza y estandariza el informe de resultados del laboratorio de PCR Covid-19.

**ACCIONES PROPUESTAS:**

1. Crear un plan estratégico de mediano y largo plazo que establezca la interoperabilidad sintáctica, semántica y organizacional como requerimiento normativo de los sistemas de información en salud.
2. Incorporar estándares internacionales (HL7 FHIR® y otros en el futuro) en el diseño y desarrollo de sistemas públicos y privados de información en salud.
3. Formar talento que permita la adopción de estándares y el diseño de sistemas de información interoperable.

**• BRECHAS EN CALIDAD DE DATOS Y SOFTWARE**

La calidad de herramientas digitales en salud es un factor crítico en diferentes dimensiones. En Chile se desarrollan instrumentos basados en estándares y buenas prácticas internacionales que permiten evaluar la calidad de software en salud. Los instrumentos incluyen la medición de madurez de soluciones innovadoras, la evaluación general de la calidad de software, incluyendo aspectos como seguridad, fiabilidad, usabilidad e interoperabilidad, más allá de la evaluación y acompañamiento en procesos de implementación de registros clínicos electrónicos, la evaluación de la calidad de datos clínicos, su uso secundario o soluciones tecnológicas específicas para telemedicina.

A pesar de la existencia, conocimiento y disponibilidad en estándares, buenas prácticas e instrumentos de evaluación en el área de calidad de software, se requiere definir, comunicar y respetar la importancia de la calidad de software y de los mecanismos y procesos existentes para implementación, revisión, chequeo y gobernanza.

**ACCIONES PROPUESTAS:**

1. Establecimiento de mecanismos de divulgación de la importancia de la calidad de software.
2. Formación de recursos complementarios de calidad de software, enfocados en el desarrollo de herramientas para el ámbito de salud digital.
3. Definición, comunicación y establecimiento de estándares, codificaciones, buenas prácticas y procesos de evaluación, para la implementación de soluciones digitales en salud.

## • DEFINICIÓN DE UN SISTEMA INTEGRADO DE DATOS

Se requiere la definición de un sistema integrado de datos y de su gobernanza, que tenga como misión ser el proveedor de datos de calidad para la investigación y toma de decisiones. Este sistema debe incluir la identificación clara de una organización central, con atribuciones claras y componentes responsables de la generación y procesamiento de datos y contextos.

Al analizar políticas de inteligencia artificial, en el contexto de la creación de la política Chilena, en diferentes países reconocen la disponibilidad de datos como un factor crítico del progreso en la materia. Algunas de estas políticas incluso elevan a los datos a la categoría de bien público, y orientan mecanismos para que actores públicos y privados los dispongan, en donde el Estado actúa como un tercero de confianza.

En general, contar con un sistema de disposición de evidencia basada en datos permite innovar tanto en políticas públicas como en el sector privado, al contar con materia prima para comprender fenómenos y resultados y habilita tecnologías emergentes, críticas para el desarrollo del país en el contexto de la Revolución Tecnológica.

La definición de un sistema integrado de datos, y de su modelo de gobernanza, requiere de la modernización del marco normativo vigente para asegurar un accionar efectivo, eficiente, transparente y confiable respecto del uso de los datos. Independientemente del modelo de gobernanza escogido, el sistema debería cumplir los principios aprobados en la Resolución N° 68/261 de la Asamblea General de las Naciones Unidas<sup>15</sup>, los que se pueden conceptualizar como imparcialidad, independencia técnica, objetividad, rectoría de la divulgación, costo-efectividad, confidencialidad, acceso o transparencia, coordinación, coherencia y cooperación.

La pandemia ha demostrado que el talento para trabajar con los datos se encuentra distribuido en la sociedad civil, la academia, el Estado y los privados. Además, se vio que la multidisciplinaria es fundamental para comprender los fenómenos complejos, como el caso de la emergencia sanitaria. El sistema integrado de datos debe ser la interfaz entre diferentes disciplinas, habilitando el trabajo colaborativo.

---

15 Disponible en <https://unstats.un.org/unsd/dnss/gp/fp-new-s.pdf>

4.

# CONCLUSIONES



## CONCLUSIONES

---

La experiencia del trabajo con datos durante la pandemia por COVID-19 en Chile ha sido desafiante y, a la vez, sumamente enriquecedora. En forma acelerada, se han coordinado actores y sistemas que previamente no dialogaban, se han identificado brechas e iniciado acciones para abordarlas, y se han abierto oportunidades para mejorar la toma de decisiones y las políticas públicas y, así, influir positivamente en la vida de los ciudadanos.

Los desafíos para un país emergente como Chile, como se ha visto, son importantes. En esta línea, se debe avanzar en un trabajo coordinado y multidisciplinario entre el sector público, la academia, la industria y la sociedad civil para reducir las brechas que afectan las distintas etapas del ciclo de datos. La pandemia forzó espacios de colaboración y la creación de soluciones innovadoras *ad-hoc*. Una vez superada la emergencia sanitaria, esta experiencia ha dejado importantes aprendizajes que podrán servir de base para políticas de largo plazo y un cambio de paradigma en la disposición y el trabajo con datos.

Finalmente, el escenario de la pandemia, la creación del Minciencia y la futura discusión constitucional generan una oportunidad importante para avanzar hacia un nuevo sistema integrado y una gobernanza de datos que potencie la colaboración y, a través de acuerdos amplios, permita fortalecer el diseño de políticas y la toma de decisiones basada en evidencia.





---

**MINISTERIO DE CIENCIA,  
TECNOLOGÍA, CONOCIMIENTO  
E INNOVACIÓN**

La Moneda s/n piso 2, ala Sur  
[contacto@minciencia.cl](mailto:contacto@minciencia.cl)

[www.minciencia.gob.cl](http://www.minciencia.gob.cl)