

! Anyone can publish on Medium per our [Políticas](#), but we don't fact-check every story. For more info about the coronavirus, see [cdc.gov](#).

Datos Abiertos para el Combate del Coronavirus

Por Marcelo Arenas, Pablo Barceló, Claudio Gutiérrez, Juan Pablo Luna, Juan Reutter, Martín Ugarte y Eduardo Undurraga



Marcelo Arenas [Follow](#)

Apr 30 · 11 min read

Desde hace años la comunidad científica viene señalando la importancia de la disponibilidad de datos e información, y las consecuencias de no tenerlos. Un caso particularmente dramático es el de las emergencias de salud pública. Existen fuertes razones para creer que las epidemias son un fenómeno sanitario que llegó para quedarse, teniendo entre sus probables causas la globalización, el movimiento de la población, la urbanización y el cambio climático [1]. Claros ejemplos de esto son las epidemias de ébola en 2014, zika en 2015 y Covid-19 en 2019 [1–4]. En este artículo describimos la preocupante situación de Chile en términos de datos para combatir la pandemia de Covid-19, y la urgente necesidad de contar con datos abiertos, desagregados y auditables.

El rol de la Mesa de Datos Covid-19 frente a los datos por paciente y su privacidad

La Mesa de Datos Covid-19 fue creada por el Ministerio de Ciencia, Tecnología, Conocimiento e Innovación para poner datos epidemiológicos y demográficos a disposición de la comunidad científica, para que ésta pudiese estudiar y caracterizar la pandemia, y aportar con antecedentes que contribuyan con información a la toma de decisiones. Una de las principales herramientas del mundo científico para esto son los modelos predictivos. Estos modelos permiten, en base a datos observados sobre el comportamiento del virus, estimar el potencial de crecimiento de la epidemia (total de

infectados, incidencia diaria, casos fatales), la demanda de recursos del sistema de salud en el tiempo (camas, ventiladores) o el riesgo de contagio para distintos grupos en la población. Y aún más, nos permiten entender cómo cambian todas esas estimaciones en distintos escenarios, con intervenciones focalizadas, globales o intermitentes. Estos modelos son muy relevantes para el debate médico-científico abierto (debido a la incertidumbre existente sobre el virus), y un insumo esencial en la toma de decisiones por parte de autoridades sanitarias y políticas.

La calidad y utilidad de los resultados de estos modelos predictivos, depende de la calidad de los datos con que se alimentan. Es natural pensar que mientras más detallados (o “desagregados”) vienen los datos, mejor va a ser la información que nos entregan para tomar decisiones. Tener más información y con un mayor nivel de granularidad permite tomar en cuenta variables que son fundamentales para generar buenas predicciones. Sabemos, por ejemplo, que un paciente Covid-19 puede contagiar a otros al menos desde el momento en el que presenta síntomas, y tal vez un poco antes. También sabemos que el tiempo que transcurre entre el inicio de los síntomas y el momento en que se confirma el diagnóstico es variable, y depende de factores tales como la disposición de un paciente a pedir ayuda médica, el tipo de seguro de salud con el que cuenta, la capacidad de diagnóstico en su sector de residencia y el nivel de saturación de los laboratorios. Todas estas variables pueden ser tomadas en cuenta para generar modelos más precisos y con mayor capacidad predictiva.

Pareciera entonces bastante obvio que debemos alimentar a estos modelos con toda la información disponible. Sin embargo, un factor muy relevante entra en conflicto: el resguardo de la identidad de los pacientes. Aunque es válido cuestionar si la privacidad de los datos debe limitar el uso que puede hacerse de ellos cuando la salud del país está en juego [5], hemos visto cómo la liberación de datos sobre pacientes Covid-19 puede llevar a formas gravísimas de discriminación tales como el apedreamiento de viviendas de pacientes Covid-19, y las amenazas de muertes y eventuales asesinatos de pacientes y sospechosos de estar contagiados con Covid-19 [6–9].

La pregunta clave es, entonces, ¿con la información que se necesita para mejorar los modelos predictivos, se puede identificar individualmente a un paciente Covid-19? Esta pregunta fue fundamental para delimitar el objetivo de la Mesa de Datos Covid-19. Inicialmente, se debía definir de forma precisa el nivel de detalle necesario para que los

modelos pudieran ser utilizados. Luego, habría que estudiar la forma en la que estos datos podrían ser disponibilizados a las universidades, centros de investigación y el público en general resguardando la identidad de los pacientes.

Para ejemplificar nuestra discusión, consideremos la información que ha publicado el Ministerio de Salud con respecto a la fecha de primeros síntomas. Específicamente, se ha publicado a nivel comunal el número de pacientes confirmados que tuvieron sus primeros síntomas en cada semana epidemiológica [10] (semana desde que la pandemia llegó a Chile). Este número por sí solo no permite saber desde cuándo un paciente está contagiando a otros, información vital para modelar una epidemia que evoluciona tan rápidamente, y que genera un número considerable de infecciones asintomáticas. Más aún, los datos publicados tratan de manera homogénea a la población de una comuna, y no permiten ver cuánto pueden variar en ella los periodos entre los primeros síntomas, la fecha de consulta y la fecha de reporte oficial. Esta variabilidad es especialmente sensible ya que podría reflejar la necesidad de reforzar distintas partes de la infraestructura montada para controlar la epidemia; por ejemplo, zonas de una comuna que requieren mayor accesibilidad o apoyo sanitario, o donde probablemente se requiere más testeo.

Si consideramos las fechas mencionadas anteriormente (inicio de síntomas, primera consulta y reporte oficial), es evidente que no contienen datos sensibles que permitirían identificar a un paciente. Más aún, el conocimiento de estas tres fechas y de la comuna en la que reside una persona no permite, bajo suposiciones razonables, identificar a sujetos particulares, lo que técnicamente se conoce como “desanonimizar”. Vale decir, se podría publicar sin riesgo estas tres fechas y la comuna donde vive un paciente, tal como lo hacen hoy los gobiernos de México [11] y Colombia [12]. De hecho, estos dos gobiernos han decidido publicar datos detallados por paciente, eliminando sólo la información personal para evitar que sean identificados directamente.

Teniendo en consideración factores como los mencionados arriba, y tomando una postura técnica, la Mesa de Datos Covid-19 identificó la necesidad de contar con una herramienta que permitiera anonimizar de manera automática los datos sensibles sobre pacientes y casos sospechosos de Covid-19. Este sistema fue desarrollado por el Instituto Milenio Fundamentos de los Datos [13], y actualmente se encuentra disponible para ser utilizado por la autoridad sanitaria, sólo requiriendo para ello contar en forma periódica

con información proveniente de los datos epidemiológicos del sistema de salud. Cabe destacar que este sistema es capaz de generar de manera automática conjuntos de datos con distintos niveles de anonimización, es decir, con mayor o menor cantidad de antecedentes. En particular, esta herramienta permite publicar conjuntos de datos similares a los que publican los Gobiernos de México y Colombia, tomando en cuenta características demográficas tales como la población y el número de contagiados por comuna para asegurar que la privacidad de las personas no se vea afectada.

La urgencia de datos abiertos, desagregados y auditables

En el contexto de las emergencias de salud pública es bien sabido que la disponibilización de datos en forma abierta, desagregados, en formatos adecuados, y en tiempo real, permite e incentiva fuertemente [14,15]:

- Que miles de ojos y mentes trabajen sobre los datos, generando comparaciones entre modelos y mayor certidumbre sobre las mejores opciones.
- La exploración y desarrollo de tópicos no vistos por los generadores de los datos, produciendo innovaciones que grupos aislados usualmente no ven.
- La auditabilidad de los procesos y decisiones tomadas, produciendo confianza en la población, y evitando discusiones y conflictos inútiles.
- Desarrollar nuevas áreas de trabajo e investigación científica no consideradas previamente sobre el devenir futuro de la crisis.
- Facilitar investigaciones cruzadas que ayudan al combate de la pandemia en otras latitudes y en los países vecinos.
- Generar ecosistemas de innovación y desarrollo de nuevos productos en el nuevo escenario actual, y en los escenarios posteriores a la crisis.

Es importante hacer notar que el uso abierto de datos, y los puntos anteriores, no se contraponen con la idea de que las autoridades competentes son quienes toman las decisiones y definiciones de estrategias ante una emergencia de salud pública, o de otro tipo. Lo que nos está diciendo la discusión anterior es que las decisiones y estrategias tomadas por las autoridades serán fundamentadas, públicas, auditables e informadas.

Lamentablemente, durante esta crisis ha quedado en evidencia que los principios mencionados más arriba no son prioritarios para las autoridades sanitarias. Pareciera ser que la repercusión política que podría generar el hacer ciertos procesos auditables por la ciudadanía, se antepone al objetivo fundamental de enfrentar la pandemia con las mejores herramientas de las que disponemos, llevándonos a tener a científicos, expertos y la ciudadanía sin poder contar con datos importantes. Y esto es un problema que va más allá de la autoridad sanitaria, creemos que está arraigado en nuestra falta de cultura de datos abiertos.

Tratándose de información estratégica para enfrentar una pandemia, la disponibilidad de los datos y su uso para contribuir con información a la toma de decisiones (resguardando, como se discute antes, la privacidad de los ciudadanos), es un deber ético y político de quienes los gestionan. Hemos visto cómo centros de investigación [16], organismos privados [17], empresas [18], entidades públicas [19,20], y ciudadanos particulares [21,22] han generado iniciativas de datos que podrían hacer una diferencia en la forma en que enfrentamos esta pandemia. Estas iniciativas, entre las cuales se encuentra la Mesa de Datos, han trabajado rápidamente en soluciones técnicas; pero hoy están a la espera de que exista la voluntad política de abrir la información del Ministerio de Salud.

¿Qué problemas debemos abordar ahora y en qué debemos trabajar en el futuro inmediato y a mediano plazo?

La actual pandemia ha llevado a una proliferación de iniciativas en todas partes del mundo para facilitar el acceso abierto a datos, y de esta forma motivar la investigación colaborativa entre la comunidad científica y los gobiernos [23–26]. Sin embargo, hay una gran preocupación por diversas falencias en el tratamiento de los datos en esta pandemia [27], y que hemos descrito anteriormente para el caso chileno. En particular, hay preocupación sobre los ámbitos en los que pueden ser usados los datos, la calidad de ellos, y sobre su captura y uso oportuno, pues todo esto puede incidir negativamente en los procesos de toma de decisiones. Entre los puntos más críticos están la imposibilidad de comparar datos entre distintas organizaciones y países dadas las diferentes interpretaciones que estos pueden tener, la falta de explicaciones detalladas y de transparencia sobre las metodologías de recolección y procesamiento de datos, y los sesgos en los resultados de la investigación científica que puedan venir de decisiones políticas [28].

De todos estos factores analizados, se desprende la urgente necesidad de desarrollar una infraestructura que permita capturar datos, y a partir de ellos generar conjuntos de datos integrados de calidad y que incorporen sus diversas dimensiones. Además, esta infraestructura debe incorporar las virtudes que proporciona un sistema de datos abiertos mencionadas en los puntos anteriores, y apoyar a la toma de decisiones para que estas cuenten con fundamentos y sean consistentes, auditables y transparentes. Por supuesto, todo lo anterior va a necesitar de una población educada en temas de datos, de especialistas formados en estos temas, y de una gobernanza de datos de alto nivel, profesional y transversal [29].

La pregunta que nos corresponde hacer en este contexto es qué problemas y temas debemos abordar ahora en nuestro país, y en qué debemos trabajar en los próximos meses. Con este análisis queremos dar inicio a una discusión sobre lo que nos parece urgente, sobre los temas en los que vamos a tener que trabajar en el corto plazo, para seguir haciendo frente a esta pandemia, y sobre los principios de transparencia y privacidad que pueden verse tensionados en distintas decisiones y estrategias.

En primer lugar, nos parece urgente que el Ministerio de Salud dé acceso desagregado y anonimizado a los datos sobre pacientes, sospechosos y descartados de estar contagiados por Covid-19, y a los datos sobre personas testeadas, positivos y negativos. También es importante contar con información actualizada sobre infraestructura hospitalaria y personal de salud. Como mencionamos antes, la transparencia sobre estos y otros datos, que incluye su publicación actualizada y responsable [30,31], y la posibilidad de combinarlos con información recogida por otras agencias sanitarias, organismos de gobierno como el Ministerio de Transportes y Telecomunicaciones, y entidades privadas en temas de comunicaciones y abastecimiento, nos va a permitir generar información relevante y de calidad para informar la toma de decisiones sobre estrategias de mitigación y optimización de recursos. Además, el uso de datos abiertos va a permitir que estas decisiones sean fundamentadas, públicas, auditables e informadas.

En segundo lugar, en los próximos meses será necesario trabajar sobre el problema de trazabilidad y, en particular, el seguimiento de personas para identificar trayectorias de transmisión que son necesarias para un período de transición a la normalidad. Será necesario entonces considerar no sólo datos que se refieren a asuntos sanitarios, sino

que también educacionales, socio-económicos y culturales. Aquí hay que tomar en cuenta políticas y regulaciones para ver cómo incorporar datos relevantes que hoy son manejados por empresas de telecomunicaciones y de redes sociales. En este punto el derecho a la privacidad de las personas juega un rol clave, lo cual se debe traducir en una recolección y uso responsable de datos [32]. Vemos aquí una necesidad urgente de tener una discusión a nivel país sobre la tensión que se va a producir entre privacidad y transparencia en el periodo de transición a la normalidad [5].

En estas dimensiones, la pandemia sólo ha venido a sincerar una realidad que comenzaba a hacerse presente. La revolución digital ha convertido los datos en un recurso esencial para el desarrollo de las sociedades. Los datos son en la práctica la copia virtual de los fenómenos observables, es la réplica virtual del mundo material. Los seres humanos acostumbrados por siglos a lidiar con el mundo material, tenemos la tendencia a traspasar esas prácticas al mundo virtual. La pandemia nos ha venido a mostrar que necesitamos una manera radicalmente diferente de abordar la revolución digital y el uso de los datos. Es lo que hemos querido mostrar aquí.

Referencias

1. Bill Gates. Pandemic I: The First Modern Pandemic. [Link](#)
2. Fighting Ebola with Information: Learning from the Use of Data, Information, and Digital Technologies in the West Africa Ebola Outbreak Response. [Link](#)
3. Press release, 1 February 2016. Statement on data sharing in public health emergencies. [Link](#)
4. Press release, 31 January 2020. Sharing research data and findings relevant to the novel coronavirus (Covid-19) outbreak. [Link](#)
5. Harald Beyer y Loreto Cox. Correr riesgos en privacidad: una conversación necesaria. 27 de Abril 2020. [Link](#)
6. La Tercera, 23 de abril de 2020. “José Tomás Vicuña, director del Servicio Jesuita a Migrantes y brote de Covid-19 en cité: Acá el gran problema no son los haitianos, ha sido el racismo”. [Link](#)

7. La República, 17 de marzo de 2020. “En Colombia vecinos atacan a paciente con Covid-19 en Municipio de Soacha”. [Link](#)
8. The India Today, 24 de abril de 2020, “Maharashtra man dies after being attacked on suspicion of being Covid-19 patient”. [Link](#)
9. The Globe and Mail, 31 de marzo de 2020. “Online harassment and death threats: The fallout from a Newfoundland funeral pits neighbours against each other.” [Link](#)
10. Informes Epidemiológicos, Ministerio de Salud. [Link](#)
11. Información referente a casos COVID-19 en México. [Link](#)
12. Casos positivos de COVID-19 en Colombia. [Link](#)
13. Instituto Milenio Fundamentos de los Datos. IMFD desarrolla sistema que anonimiza de manera automática datos de pacientes COVID-19. [Link](#)
14. European Community. What is open data? [Link](#)
15. Open Knowledge Foundation. Open Data Handbook. [Link](#)
16. Webinario “Desafíos para modelar y predecir la Epidemia de COVID-19 en Chile”, 16 y 17 de abril. Primer y segundo video de la conferencia.
17. Banco Mundial. Understanding the Coronavirus (COVID-19) pandemic through data. [Link](#)
18. Revista Capital. Emprendedores que se reinventan en cuarentena, 31 de marzo de 2020. [Link](#)
19. Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. Datos COVID-19. [Link](#)
20. División de Gobierno Digital, Ministerio Secretaría General de la Presidencia. CoronApp: La nueva aplicación de Chile para combatir la pandemia. [Link](#)
21. Jorge Pérez. Datos COVID19 Chile. [Link](#)
22. Panel Interactivo COVID-19 en Chile. [Link](#)

23. Open data, data use, and Covid-19. [Link](#)
24. Virus Outbreak Data Network (VODAN). [Link](#)
25. COVID-19 Data Portal. [Link](#)
26. Utrecht University. Making data on Covid-19 available to the scientific community. [Link](#)
27. Evgeny Bobrov. Open Data can be advanced by the Covid-19 pandemic, but will still require a comprehensive approach. [Link](#)
28. International Science Council. A Data Ecosystem to Defeat Covid-19. [Link](#)
29. Mark D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, artículo 160018 (2016). [Link](#)
30. Data Responsibility in the Covid-19 Response. [Link](#)
31. Marcello Lenca y Effy Vayena, On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature Medicine* 26, páginas 463–464, 2020. [Link](#)
32. Council of Europe. Covid-19 and Data Protection. [Link](#)

[Datos Abiertos](#)[Coronavirus](#)[Chile](#)[Transparencia](#)[Privacidad](#)[About](#) [Help](#) [Legal](#)

Get the Medium app

