

COVID19 y datos abiertos

Claudio Gutierrez

Universidad de Chile, Chile &
Instituto Milenio Fundamentos de los Datos

OLinFER / CEPAL, Biblioteca Hernán Santa Cruz
18 de Noviembre de 2020

un tema con precedentes (I)

Desde hace varios años que la comunidad científica viene señalando la importancia de la disponibilidad de información y datos --y las consecuencias de no hacerlo-- en las emergencias de salud pública:

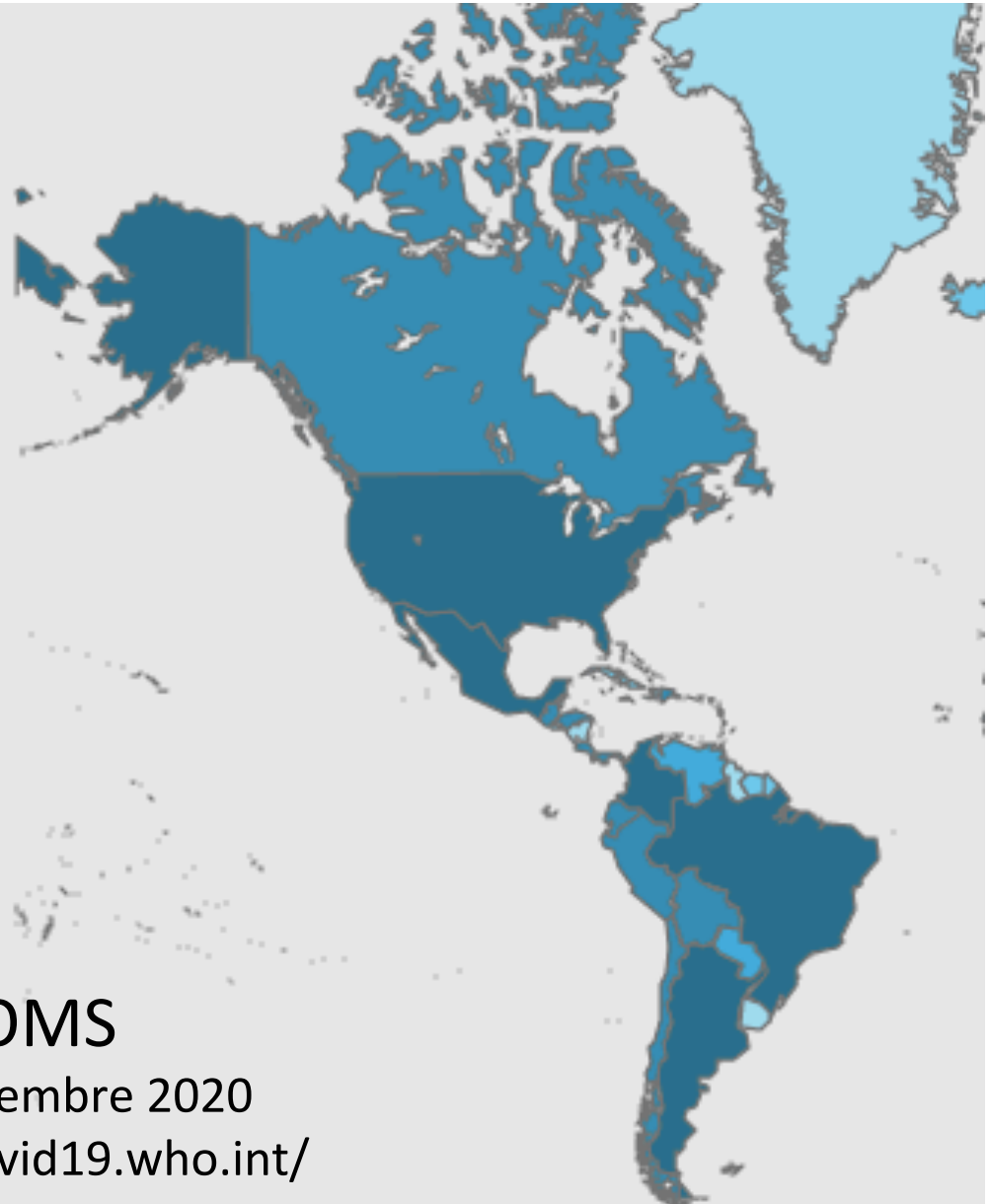
- Zika (2016),
- Ebola (2018),
- Covid-19 (enero 2020)

un tema con precedentes (II)

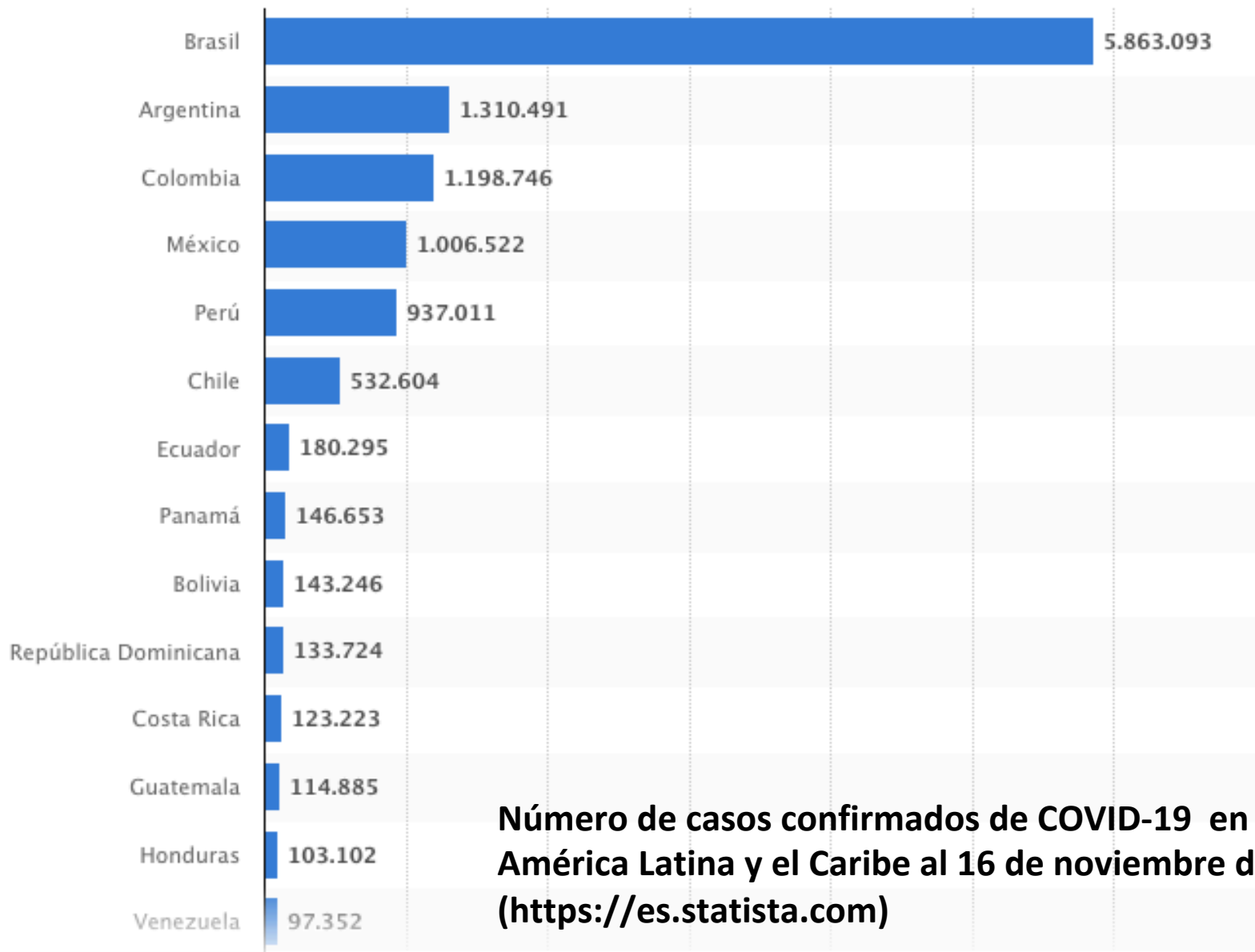
El movimiento de datos abiertos tiene raíces en diversos movimientos por la apertura en el mundo digital:

- Free software
- Open source
- Creative Commons
- Diversas áreas científicas (e.g. Medio ambiente, geofísica, etc.)
- Etc.

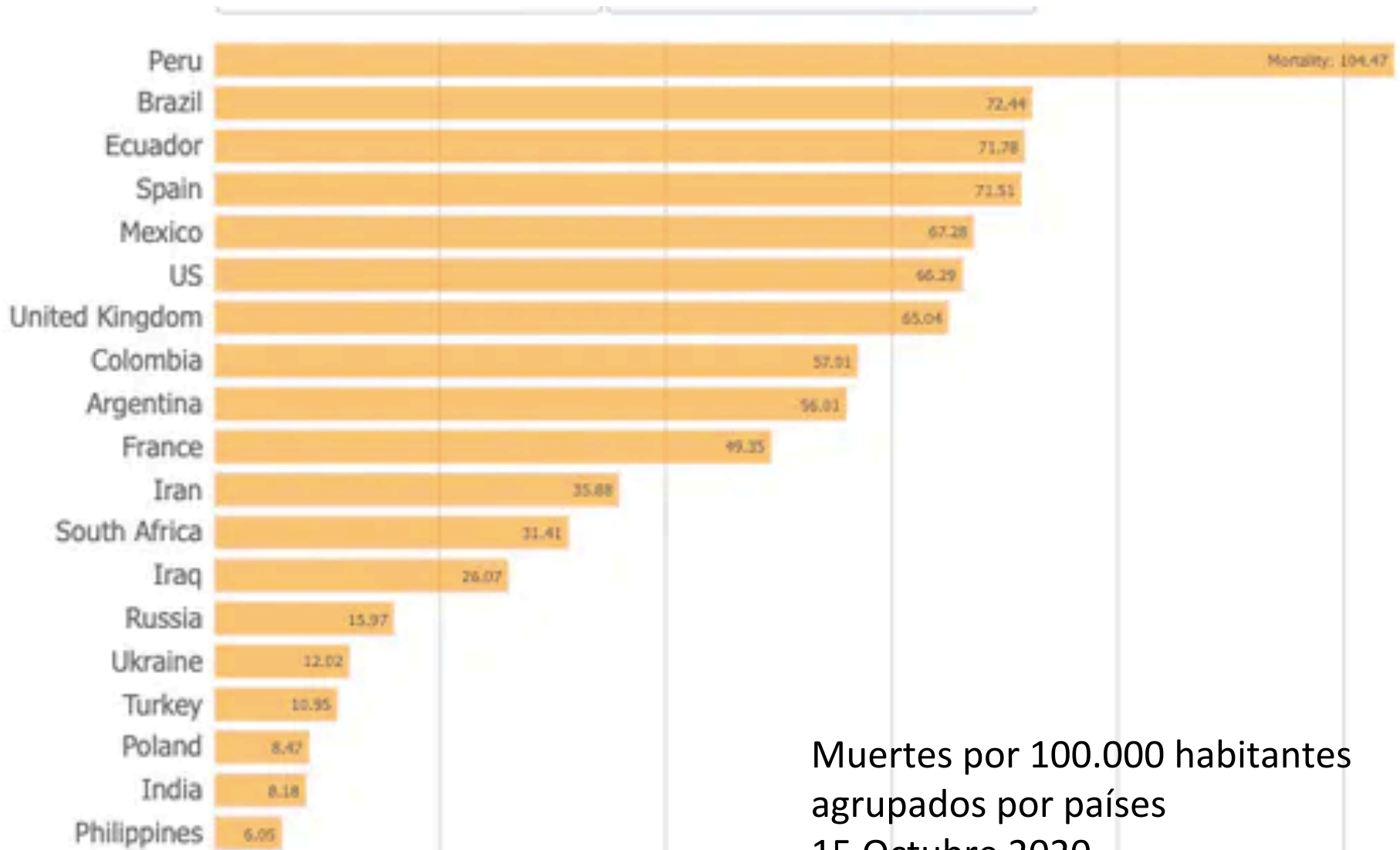
La pandemia del COVID19 ha realzado la relevancia de los datos abiertos en el área de salud



OMS
11 noviembre 2020
<https://covid19.who.int/>



Número de casos confirmados de COVID-19 en América Latina y el Caribe al 16 de noviembre de 2020 (<https://es.statista.com>)



Muertes por 100.000 habitantes
agrupados por países

15 Octubre 2020

<https://theconversation.com>

latinoamérica y COVID19

El problema principal en Latinoamérica no son los datos:

- Strategies were incomplete lacking extensive testing and tracking measures.
- Response effectiveness undermined by socioeconomic conditions and health capacities.
- A poor pandemic management (communication and coordination) influenced results.

Responses to COVID-19 in five Latin American countries. *Health Policy and Technology*, December 2020.

aunque...

“The epidemiological data quality and availability is a significant limitation that highlights the disease monitoring shortcomings in the selected countries. **This is an important caveat for a well-planned recovery strategy.**”

Responses to COVID-19 in five Latin American countries. *Health Policy and Technology*, December 2020.

casos de uso de datos abiertos durante la pandemia:

- Curation of data from diverse sources for better visualization of the scope of the pandemic and information exchange.
- Creation of mobile applications to empower consumers to make data-informed decisions on how to adjust their retail behaviors to reduce their personal risk.

The Potential Role Of Open Data In Mitigating The COVID-19 Pandemic: Challenges And Opportunities. *Health Affairs*, November 2, 2020

desafíos de datos:

- de-identify, reformat, publish, and promote the use of existing government data require considerable financial and staff resources and expertise and these financial investments must be ongoing to ensure sustainability.
- data quality, timeliness, completeness, and availability. There has been under-reporting of cases, particularly where there is limited test availability.
- countries such as Korea and Singapore that disclosed granular data, additional challenges have been privacy concerns and increased social stigma, which can in turn discourage community members from being tested.
- provide sufficient context and metadata (information about the data, such as a codebook, data collection methods, coverage, limitations, and so forth) to ensure correct interpretations of the data by end users.

The Potential Role Of Open Data In Mitigating The COVID-19 Pandemic: Challenges And Opportunities. *Health Affairs*, November 2, 2020

algunas recomendaciones:

- Advocate for Governmental Transparency including release of Open Data
- Continue to invest in Open Data Infrastructure
- Enhance the quality of Open Health Data
- Ensure Data Privacy
- Cultivate the Open Data Ecosystem

desafíos para datos abiertos:

1. Desarrollar un análisis fino y estratégico del manejo de datos durante el COVID19
2. Entender qué significa “open data” más allá de una *buzzword* de moda

Algunas fuentes:

- <https://opendatawatch.com/what-is-being-said/data-in-the-time-of-covid-19/>

- <https://datascience.nih.gov/covid-19-open-access-resources>

- COVID-19 data sources in Latin America and the Caribbean. 2020. <https://doi.org/10.1016/j.tmaid.2020.101750>

Lo que me gustaría que
conversáramos

tres temas

1. Breve revisión a la noción de “datos” (y a la noción de “datos abiertos”)
2. El ciclo de los datos: las enseñanzas que deja el COVID19 en cada paso
3. El gran tema a futuro en este ámbito: los datos (y lo digital) como comunes

¿qué son los “datos abiertos”?

- COVID19 vino a instalar ese movimiento en el centro de la discusión:
 - Nociones de: datos disponibles, datos públicos, gobierno de esos datos, privacidad, propiedad, etc. han pasado a ser temas comunes
- ¿Qué se quiere expresar cuando hablamos de datos abiertos?

¿de qué “datos abiertos” se está hablando?

¿Las fuentes de datos?

- sobre personas, tests PCR, contagios, situación económica, condiciones de vivienda, infraestructura de salud, personal de salud,

¿La información difundida y comunicada?

- Estado de la pandemia, cuidados necesarios, medidas a tomar, recursos disponibles, videos informativos,

¿El conocimiento y su difusión?

- Artículos científicos, revistas, patentes, licenciamientos, conocimiento materializado: fármacos, instrumentos, materiales laboratorio, fungibles, elementos de protección personal, etc.

ecosistema abierto

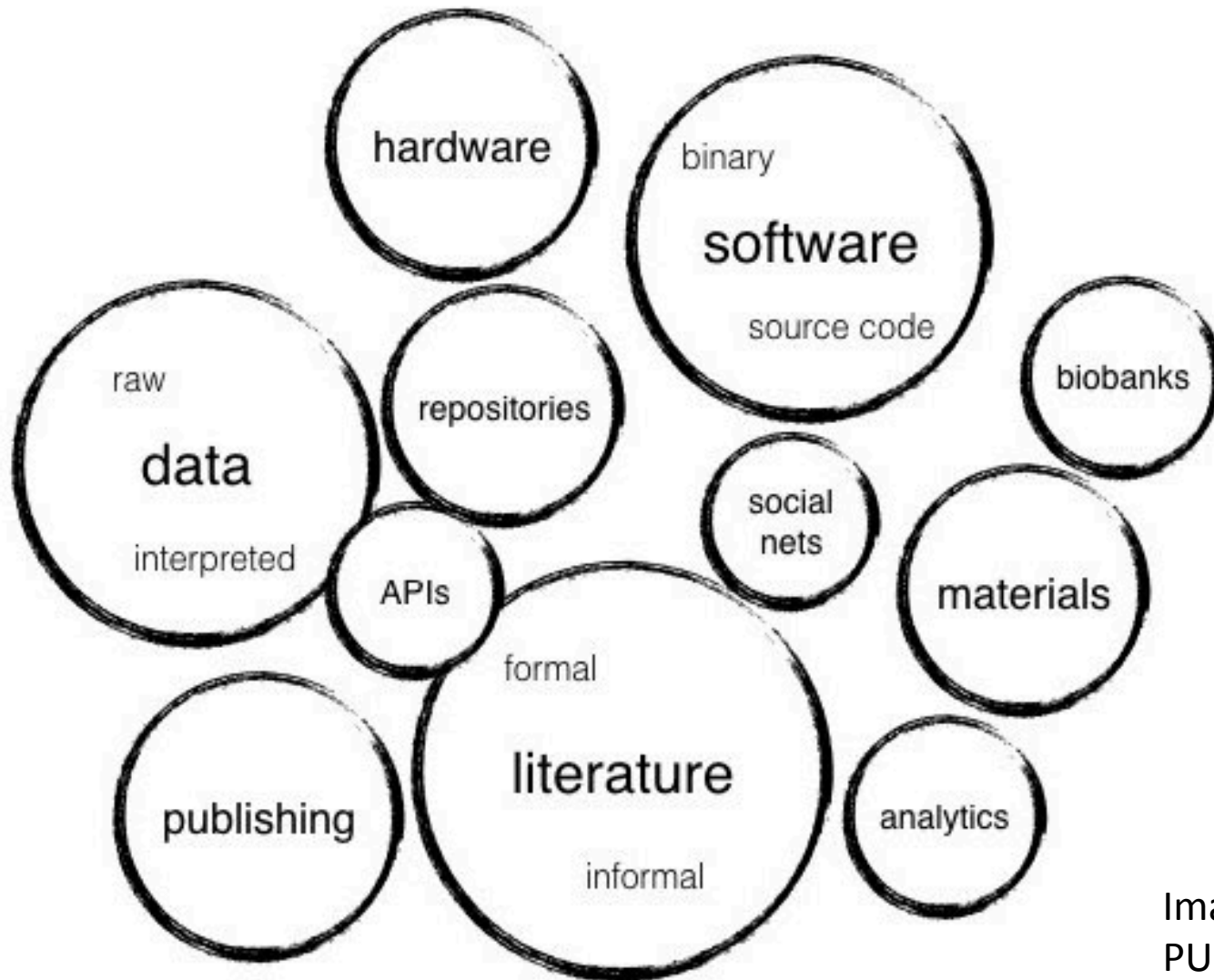


Image by
PUNEET KISHOR

importante entender las diferencias

Datos: digital, marcas mínimas, formato flexible, evidencia, precisión, replicabilidad, memoria casi infinita (Informática, Computación,...)

Información: in-formar, sentido, interpretación, lenguaje, temporalidad, contexto, humano (Periodismo, Humanidades,...)

Conocimiento: procedimiento, actividad, reglas, aprendizaje, experiencia, razonamiento, teoría, ciencia, saber, máquina (Ciencias, Automatización,..)

¿y qué es lo “abierto”?

- Lo no cerrado, no cercado, libre acceso
- Lo usable, apropiable, asimilable, consumible
- Lo replicable, transportable, procesable

¿Por qué no todo está abierto?:

Restricciones materiales:

- propiedad, preservación, depreciación, costos, escasez,...

Restricciones virtuales:

- privacidad, fiabilidad, abundancia,...

(más adelante: datos como recurso *común*)

“Apertura” en tiempos de COVID19



Nos concentraremos en “datos abiertos”

Qué dejamos fuera: información y conocimiento.

Por ejemplo:

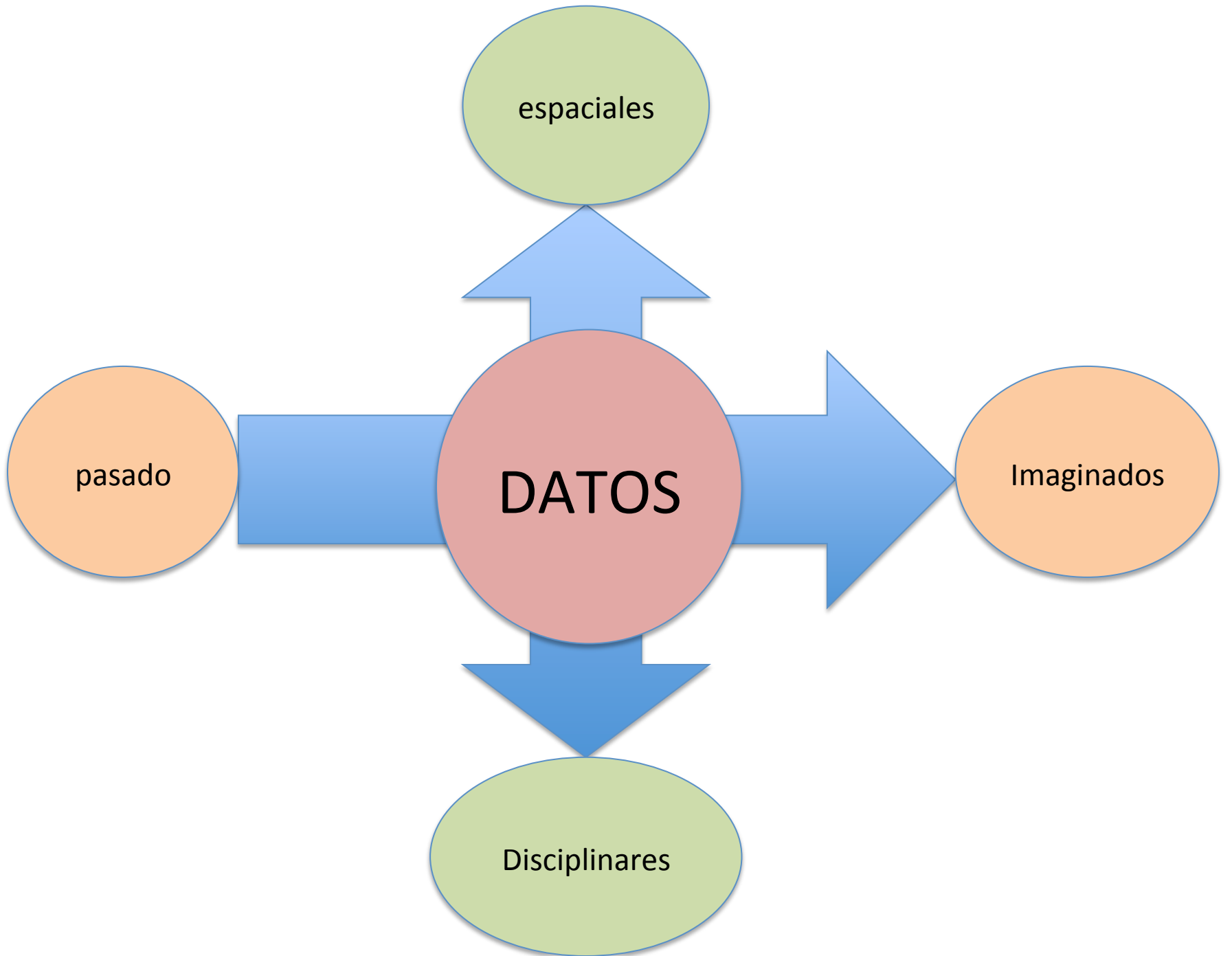
- Movimientos de apertura del conocimiento sobre *papers* sobre COVID19 en *journals*
- Movimientos de cerco del conocimiento en fármacos, tratamientos y vacunas
- Movimientos de apertura cultural de creadores en apoyo a encierro por COVID19 (música, videos, imágenes, textos)
- Movimiento de apertura de información y conocimiento de ciertas instituciones (ciertas bibliotecas, ciertas editoriales, etc.)

El ciclo de los datos

Ventajas tradicionales de los datos abiertos

La disponibilización de forma abierta de datos, desagregados, en formatos adecuados, y en tiempo real, permite e incentiva:

- a) que cientos o miles de ojos y mentes trabajen sobre los datos generando comparaciones entre modelos y certidumbre sobre las mejores opciones;
- b) exploración y desarrollo de tópicos no vistos por los generadores de los datos, produciendo innovaciones que grupos aislados usualmente no ven;
- c) auditabilidad de los procesos y decisiones tomadas, confianza en la población, evitando discusiones y conflictos inútiles;
- d) desarrollar nuevas áreas de trabajo e investigación científica no consideradas previamente sobre el devenir futuro de la crisis;
- e) facilita investigaciones cruzadas que ayudan al combate de la pandemia en otras latitudes y en los países vecinos;
- f) genera ecosistema de innovación y desarrollo de nuevas y desconocidos productos en el nuevo escenario actual y post-crisis.



espaciales

DATOS

Disciplinares

pasado

Imaginados



Recolección

- Los de personas, por su naturaleza, distribuida (ir a buscar, no esperar que lleguen)
- Por su modelo, unificado (e.g. atención pública y privada)
- Nudo crítico: digitalización y formato de los datos
- Calidad de sistemas de ingreso de datos
- Hacia sistemas automatizados de recolección
- Sistema distribuido de manejo de datos
- Interoperabilidad de datos de base
- Limpieza y calidad de ingreso de datos

Caso Chile: mecanismos “primitivos” e informales paralelos;

Transporte y comunicación

- Diferentes niveles: atención primaria hacia arriba; atención secundaria y terciaria
- Diseño de red digital de canales comunicantes de datos flexible, fiable y resiliente
- Interconexión con otras fuentes de datos (sociales, económicos, geográficos)
- Salidas hacia las unidades de análisis y tomas de decisión
- Incluye “datos” no digitales (muestras, etc.)

Caso Chile: red inexistente; lentitud; incorporación errores;

Clasificación y almacenamiento

- Ontologías y metadatos estándares y que integren factores epidemiológicos y culturales y sociales
- Modelos de interoperabilidad con datos de otras áreas (especialmente sociales y geográficos)
- Balances de centralización versus distribución
- Políticas de gobierno/gobernanza de almacenamiento, accesos, manejo
- Estándares éticos de seguridad y privacidad
- Monitoreo de calidad de datos en todos los niveles
- Participación e inclusión de personas y comunidades

Caso Chile: centralizado; artesanal; sin políticas de gobierno de datos; sin participación;

Análisis y transformación

- Distribuido, abierto y transparente
- Necesidad de fuentes históricas y comparativas (otras zonas y países)
- Diferentes modelos epidemiológicos
- Focalizado en bien común
- Códigos de ética (usos, anonimización, etc.)
- Definir derechos de propiedad intelectual

Caso Chile: opaco, privado, sin estándares éticos

Disponibilización y publicación

- Abierta y desagregada (datos crudos)
- Documentada e interoperable (metadatos)
- Transparencia activa (publicación) y pasiva (mecanismos de respuesta y disponibilización)
- Segura y ética (temas de privacidad)
- En beneficio de las comunidades y la sociedad donde fueron recolectados
- Consistente a lo largo del tiempo

Caso Chile: agregados; inconsistentes; informes opacos, incompletos, manipulados;

Datos Abiertos para el Combate del Coronavirus. 30 Abril 2020.

<https://medium.com/@marcelo.arenas/>

Archivo

- Decisiones basadas en políticas y principios de ética, no económicos y de eficiencia
- ¿Qué descartar / que archivar?
- ¿Dónde quedan los archivos?
- ¿Quién queda a cargo?
- ¿Cómo se enlaza con los datos activos?

Caso Chile: sin políticas de archivo

Datos como comunes

Comunes

- Entendemos por "comunes" recursos compartidos por una comunidad. No son directamente públicos ni privados.
- No son públicos, en el sentido restringido de ser de libre acceso o goce; son mucho más que eso; son creados y recreados por todos.
- No son privados, lo que no significa que no sean de nadie, ni que esos recursos surjan solos, ni que se mantengan solos.

debilidades de lo abierto

La noción de (datos) “abiertos” da cabida a algunos equívocos:

- separar funciones de acceso de las de creación, mantenimiento y uso
- mantenerlos ligados a la dicotomía de propiedad privada versus pública
- desapegarlos de su característica de creación fluida, temporal, y verlos como objetos estáticos
- esconder/olvidar la responsabilidad de las comunidades (y dejarlo todo a una poder o institución ajena)

comunes

- Abiertos a todos (disponibilidad y creación)
 - caso datos: en todo el ciclo y dimensiones
- Control por las personas y la comunidad
- Conciencia de cuidado de una infraestructura
- Normas de funcionamiento democrático
- Instituciones que encarnen el modelo
- Confianza y reciprocidad

a modo de conclusión:

- + Los datos son un común
- + La salud es un común (*)

= los datos de salud son un (doble) común

(*) La salud, entendida como el bienestar físico, mental y social es un recurso generado, construido y mantenido por todos los integrantes de una comunidad

gracias x la atención

Claudio Gutierrez
cgutierrez@dcc.uchile.cl