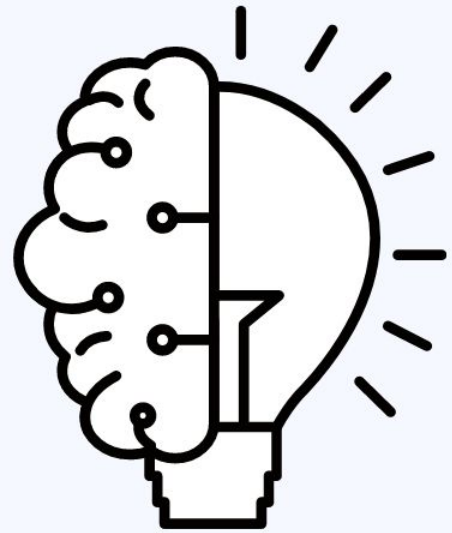


Ciencia de datos

Para el sector público de salud



Análisis de prefactibilidad y valor público

En colaboración con...



@RLadiesConce



MÓDULO 1: Nivelación y conceptos básicos

Actividades sincrónicas (2 hrs cada uno)

Fecha	Hora	Tema
09/12/2020	18.45h	Aspectos generales del curso
15/12/2020	18.45h	Introducción a R y RStudio
17/12/2020	18.45h	Estructura de datos y operadores
22/12/2020	18.45h	Importación de datos
29/12/2020	18.45h	Análisis prefactibilidad y valor público

28 Diciembre 2020

Submesa de Datos COVID-19 presenta análisis de su ejercicio de disposición de información para enfrentar la pandemia

Comparte:



La instancia creada al alero del Ministerio de Ciencia para facilitar el acceso a información de la emergencia sanitaria, elaboró el documento “Una mirada a la cultura de datos en Chile” que reúne el análisis de esta experiencia, lo aprendido y los desafíos para dar continuidad al trabajo iniciado con datos para enfrentar la pandemia más allá de la emergencia.



Una mirada a la cultura de datos en Chile

DESAFÍOS Y LECCIONES DE LA
PANDEMIA POR COVID-19

DICIEMBRE 2020



Criterios de calidad de datos para el manejo de la salud digital



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", Tesis grado CIMT Universidad de Chile https://cimt.uchile.cl/wp-content/uploads/2020/10/20201027_Tesis_Laura_Araya.pdf
Adaptado de C. Batini, C. Cappiello, C. Francalanci, A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Computing Surveys (CSUR)*, vol. 41, p. 16, 2009 <http://dimacs-algorithmic-mdm.wdfiles.com/local--files/start/Methodologies%20for%20Data%20Quality%20Assessment%20and%20Improvement.pdf>

Perspective | [Free Access](#)

Population Health in the Time of COVID-19: Confirmations and Revelations

ANA V. DIEZ ROUX 

First published: 18 August 2020 | <https://doi.org/10.1111/1468-0009.12474> | Citations: 1

 SECTIONS

 PDF  TOOLS  SHARE

THIS HAS BEEN AN INCREDIBLE PERIOD FOR PUBLIC HEALTH: A time of unprecedented discussion of the science of population health in the media and in society, and a time of uncertainty but also of dramatic social action to protect health in ways that none of us would have predicted just a few months ago. Many aspects of the pandemic we are living through are linked to specific characteristics of the SARS-CoV-2 virus: how it is transmitted to and infects individuals, what



[Volume 98, Issue 3](#)

September 2020

Pages 629-640

This article also appears in:
[Population Health](#)

Advertisement



Recurso recomendado:

(Esta presentación se basa bastante en ello)

Curso “Formulación de proyectos de ciencia de datos para directivos públicos”

UAI - Laboratorio de Gobierno Chile

<https://innovadorespublicos.cl/cases/37/>

<https://gobierno.uai.cl/centros/goblab/proyecto-ciencia-de-datos-para-directivos-publicos/>

¿Prefactibilidad?

Condiciones que favorecen (o no) el desarrollo de un proyecto de CD.

¿Agrega valor?



**Idea de agregación
de valor**



**Suficiente apoyo
político**



**Factibilidad técnica
de ejecución**

Más información sobre el modelo en <http://opensaludlab.blogspot.com/2019/01/framework-para-el-analisis-de.html>

Análisis de prefactibilidad

1. Es un problema real y que genera valor (significativo)
2. Hay capacidad de abordar (o actuar) el problema
3. Es prioritario para la institución y hay compromiso (de directivos)
4. Hay datos (y acceso) a datos relevantes
5. Hemos identificado los riesgos

Lo más probable es que falles

... o que el proyecto quede a medias

Análisis de prefactibilidad

1. **Es un problema real y que genera valor (significativo)**
2. **Hay capacidad de abordar (o actuar) el problema**
3. Es prioritario para la institución y hay compromiso (de directivos)
4. Hay datos (y acceso) a datos relevantes
5. **Hemos identificado los riesgos**

Uso responsable y ético de los datos

Problema real y significativo

1. ¿Cuál es la necesidad?
2. Dimensiones del problema:
 - ¿Quiénes/qué se ve afectado por este problema?
 - ¿Cuántas personas/instituciones/poblaciones se ven afectadas y en qué medida?
3. ¿Por qué resolver este problema es una prioridad para tu organización?
4. ¿Cuál es la solución actual? (Si la hay)
5. ¿Hay alguna solución no basada en datos que sería más factible?

Entender profundamente el problema (ver videos OSL)

Diseño centrado en el usuario

Problema real y significativo

¿Por qué resolver este problema es una prioridad para la organización?

Definir los objetivos...

- La solución técnica (p. ej., modelo predictivo) no es el objetivo
- El objetivo debe ser medible
- Alcanzar el objetivo debería ayudar a solucionar el problema
- Algunos objetivos habituales son mejorar/maximizar/aumentar o disminuir/mitigar/reducir un resultado o una métrica
- Las limitaciones típicas incluyen el presupuesto, la falta de capital humano, las restricciones legales, la voluntad política y la licencia social
- Considera las compensaciones (*trade-offs*) entre objetivos contrapuestos

Capacidad de actuar

1. Es parte de nuestra facultad legal
2. Hay un proceso/actividad/servicio concreta que aborde el tema
3. Están los recursos adecuados:
 - Presupuesto
 - Humanos (externalización? asesoría?)
 - Tecnología
 - Liderazgo (político y técnico)

Capacidad de formación interna vs compra externa

Compromiso de recursos

Acciones

- Las acciones son las actividades que las instituciones hacen/pueden hacer respecto de un problema. Implica la asignación de recursos, por ejemplo, inspeccionar instalaciones, brindar servicios preventivos, promoción, entrega de prestaciones, consultas, etc.
- En teoría, las acciones podrán ser realizadas de mejor forma cuando la institución tenga la información generada en el proyecto de ciencia de datos.
- Las acciones ideales deberían ayudarte a alcanzar los objetivos definidos anteriormente.

Es una prioridad (el problema)

1. ¿Ha sido manifestado por escrito en un plan existente?
2. ¿Los directivos han manifestado su apoyo?
3. ¿Hay un sponsor?
4. ¿Hay consenso de que es necesario abordar el problema?
5. Existe compromiso de:
 - Modificar las acciones o cómo se toman las decisiones a consecuencia de los análisis
 - Validar la solución (medir el impacto)

Si no es prioridad o no hay apoyo, deja el proyecto.

Datos

1. ¿Hay datos y se pueden acceder a ellos?
2. ¿Son suficiente útiles para hacer algo?
3. ¿Cuál es la cultura organizacional respecto al manejo de datos?

Matriz de madurez de los datos

Center of Data Science and Public Policy, University of Chicago

<http://www.datasciencepublicpolicy.org/home/resources/datamaturity/>

<https://innovadorespublicos.cl/uploads/matriz-de-madurez-de-datos-3.docx> (versión español UAI)

Data Maturity Framework

Organizational Readiness Scorecard

Area	Lagging	Basic	Advanced	Leading
Staff Buy In	Staff at the organization have some idea that data exists but doesn't understand it is important	There are a few individuals who deeply understand the data available and what can be done with it	Organization has a clear idea of how data can be used to drive business decisions beyond justification of funding	Organization has a culture of data within the organization and demands data to justify all programmatic decisions
Data Collector Buy In	On the ground staff provide data seldomly, sporadically, or incompletely because they are required to but it is seen as a hindrance to their "real job"	On the ground staff regularly provide data because they are required to	On the ground staff provide data on a regular basis and eventually get actionable insights in return	On the ground staff provide data in real time and make decisions based on the data and insights available to them, and offer suggestions on what is collected/what information they could use to improve their job effectiveness
Leadership Buy In	Leaders at this level fundamentally don't know how data can help advance the organization's mission.	Leadership wants to use data but don't have a clear path forward to use data	Leadership has a clear idea of how data can be used to drive business decisions beyond justification of funding	Leadership builds a culture of data within the organization and demands data to justify all programmatic decisions
People Resources	Individual stakeholders maintain siloed data sets	The organization knows how data can help, what data they need, and are able to access it, but lack the in-house data skills, tools, or infrastructure to be able to turn data into meaningful insights that affect human action.	Organizations know how data can help, what data they need, and are able to access it, but lack either the infrastructure or the people to be able to turn data into meaningful insights that affect human action.	The organization has dedicated staff who own data storage AND data content owners who own the cleaning and rigor of the data
Data Use Policy	No policies exist around use, transfer, and sharing of data	Organization has policies in place for the use, transfer, and sharing of data but it does not cover all data that exists within the organization	Organization has policies in place for the use, transfer, and sharing of data internally	Organization has policies in place for the use, transfer, and sharing of data internally and externally
Intervenor Buy In	No partnerships exist	Partnerships exist but data is not shared	Partnerships exist and have policies and technology in place to share data occasionally or through a manual process	Partnerships exist and have policies and technology in place to share data in real-time
Funder Buy In	Funders do not require data other than vanity metrics	Funders ask for key performance metrics	Funders ask for key performance metrics and provide funding for data infrastructure and maintenance	Funders require data driven decision making and provide funding for data infrastructure, maintenance, and usage

Data Maturity Framework

Data and Tech Readiness Scorecard

Category	Area	Lagging	Basic	Advanced	Leading
How is Data Stored	Accessibility	Only accessible within the application where it is collected	Can be accessible outside the application but proprietary format, requiring specialized analysis software	All machine readable in standard open format (CSV, JSON, XML, database)	All machine readable in standard open format and available through an API
	Storage	Paper	PDFs or Images	Text Files	Databases
	Integration	Data sits in the source systems	Data is exported occasionally and integrated in ad hoc manner	Central data warehouse - realtime aggregation and linking (Automatic)	External data also integrated
What is Collected?	Relevance and Sufficiency	The data you are collecting on subjects of interest is irrelevant to the problem you want to solve: ie you want to do predict which students need extra support to graduate on-time but don't have data on graduation outcomes	Some of the data you have is relevant, but it is insufficient because key fields are missing, ie no data on academic behavior or attendance history, etc.	You have data that is helpful and relevant for solving the problem but not sufficient to solve it well. ie you have yearly academic and demographic information but are missing extra-curricular activities, or interventions they were targeted with	You have all the relevant data about all the entities being analyzed and it's sufficient to solve the problem you are tackling
	Quality	Missing rows (people/address level entities missing in the data)	Missing columns (variables missing)	No missing data but errors in data collection such as typos	No missing data and no errors in data collection
	Collection Frequency	Once and never again	yearly	frequently	realtime
	Granularity	City level aggregates	Zipcode/Block level aggregates	Individual level (person or address) level data	Incident/Event level data
	History	No History Kept - old data is deleted	Historical data is stored but updates overwrite existing data	Historical data is stored and new data gets appended with timestamp, preserving old values	All history is kept and new data schema gets mapped to old schema so older data can be used
Other	Privacy	No privacy policy in place	no PII can be used for anything	ad-hoc approval process in place that allows selected PII data to be used for selected/approved projects	Software defined/controlled privacy protection that allows analytics to be done while preserving privacy based on predefined policies
	Documentation	no digital documentation or metadata: data exists but field descriptions or coded variables are not documented	data dictionary exists (variables and categories defined)	data dictionary plus full metadata available (including conditions under which the data were captured)	data dictionary plus full metadata available including collection assumptions, what's not collected, and potential biases

Madurez de procesos



Harvard Business Review
América Latina

La auditoría de proceso

por Michael Hammer

Abril 2007

Reimpresión R0704H-E

Riesgos

Reputacionales (**éticos**, privacidad, seguridad, legales)

Licencia social

No funcionará lo suficientemente bien

Nunca será implementado / usado

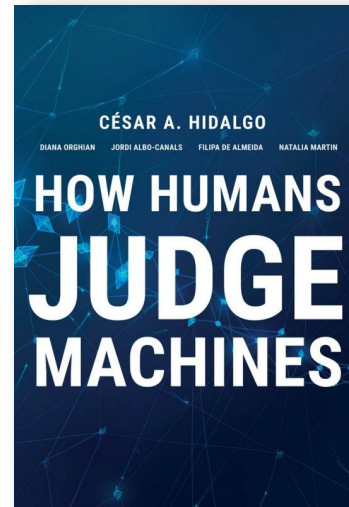
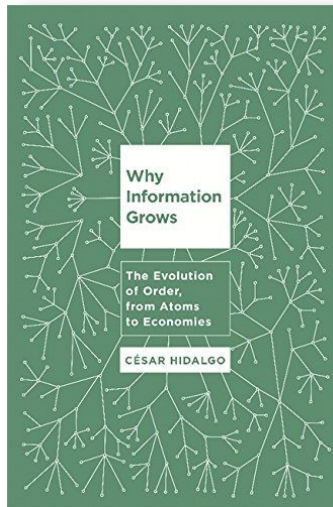
¿Impacto en los *stakeholders*? El sector público es un sistema (muy) complejo*

* Sistemas complejos y gestión pública, Mario Waissbluth. DII Universidad de Chile

<http://www.dii.uchile.cl/~ceges/publicaciones/99%20ceges%20MW.pdf>

Recomendaciones

César Hidalgo



Encuesta ciudadana sobre política nacional de inteligencia artificial

<https://www.minciencia.gob.cl/politicalA>

¿Para qué?

Objetivos

Análisis

Datos

Acciones

“Para tomar mejores decisiones...”

¿Qué decisiones?
¿Qué es una mejor decisión?
¿Por qué piensas que se están
tomando malas decisiones?
¿Qué cambiaría si decidieras distinto?

“Para ser más eficiente...”

¿Qué es ser eficiente?
¿Para qué quieres ser eficiente?
¿En qué medida quieres mejorar?