



CAMBIO DE ROLES 2022

DATOS PERSONALES
INFORME TÉCNICO



OBSERVATORIO
DE FAMILIAS Y JUVENTUDES



DIPUTADOS
ARGENTINA

INDICE

- **DATOS PERSONALES** 5
Edgardo Martínez - Especialista del BID
- **INTELIGENCIA ARTIFICIAL** 9
Dr. Ricardo Rodríguez - ICC - UBA - CONICET
- **PROTECCIÓN DE DATOS** 13
Dr. Daniel Yankelevich - Director Organizador Fundar
- **GOBERNANZA DE DATOS Y ADMINISTRACIÓN** 23
Lic. Cecilia Ruz - Dirección Analítica de Datos de AFIP
- **USO DE DATOS DE SALUD** 29
Dra. Verónica Xhardez - Investigadora de Big Data
- **IMPACTO REGULATORIO** 33
Delia Rodrigo - Especialista Senior en Mejora
Regulatoria - BID



DATOS PERSONALES

**EDGARDO MARTÍNEZ
ESPECIALISTA DEL BID**

¿QUE SON LOS DATOS PERSONALES?

Es aquella información que sirve para identificar de forma directa o indirecta a una persona, es decir, indica de qué persona se trata (como el nombre, dirección o número de identificación), o da pistas sobre la identidad de la misma

Hay 12 países con leyes de protección de datos personales vigentes Argentina, Bahamas, Brasil, Colombia, Costa Rica, Ecuador, México (con dos leyes), Nicaragua, Panamá, Perú, Trinidad y Tobago y Uruguay.

DESCONOCIMIENTO SOBRE LO QUE SON LOS DATOS PERSONALES

El 51% de las personas que respondieron la pregunta de “¿Le parece que compartir sus datos personales tiene más beneficios que riesgos?” con “Más riesgos que beneficios” (eje de la izquierda), estuvieron de acuerdo con la pregunta “¿El gobierno debería multar a las personas que son coronavirus positivo y no permiten la geolocalización de sus celulares?” (eje de la derecha). Considerar que es riesgoso compartir los datos personales pero facultar al gobierno para multar a quienes no permiten la geolocalización podría representar una discordancia y reflejar un desconocimiento de los datos personales.

ALGUNAS PERSONAS NO LE HALLAN VALOR A SUS DATOS PERSONALES

El 60% de quienes contestaron No a la pregunta “¿Sabe lo que hacen los gobiernos con sus datos personales?” contestaron Sí a la pregunta “¿El gobierno puede proporcionar beneficios al recopilar sus datos y el de otras personas para

superar los riesgos potenciales en épocas de pandemia?”. La voluntad de entregar esta información al Estado sin la necesidad de saber qué se hace con esta, podría reflejar que estas personas no le hallan un valor a sus datos personales.

Ejes de los estándares de protección de datos para los estados iberoamericanos

- Principios y derechos.
- Régimen de transferencias internacionales de datos.
- Autoridad administrativa, independiente, especializada y estable.
- Aplicación territorial de las leyes en función de la persona.

Reglamento General de Protección de Datos de la Unión Europea

La legislación europea tiene una marcada influencia en la legislación latinoamericana y del Caribe. Es posible clasificar las leyes entre las previas y las posteriores al Reglamento

- Perspectiva desde la que se habilita el tratamiento de los datos.
- Alcances de su ámbito de aplicación territorial.
- Cambio de perspectiva de leyes reactivas a leyes preventivas.

<https://www.datasketch.co/bid/datos-personales-y-leyes/>



INTELIGENCIA ARTIFICIAL

DR. RICARDO RODRÍGUEZ
ICC - UBA - CONICET

INTELIGENCIA ARTIFICIAL

“La inteligencia artificial (IA) refiere a sistemas que muestran un comportamiento inteligente al analizar su entorno y tomar acciones, con cierto grado de autonomía, para lograr los objetivos que se plantea” Unión Europea 2019.

ENFOQUES DE IA

- Desde el Conocimiento
- Desde los Datos



<https://santiferris.com/inteligencia-artificial/cuatro-aproximaciones-a-la-inteligencia-artificial/>

- Pensar de forma humana: “realizar actividades asociadas con pensar de manera humana, como la toma de decisiones, la solución de problemas o el aprendizaje” ([Bellman](#), 1978).
- Actuar de forma humana: “la realización de funciones que requieren inteligencia cuando las realizan personas” ([Kurzweil](#), 1990).

- Pensar de forma racional: “El estudio de los cálculos que hace posible percibir, razonar y actuar” (Winston, 1992)
- Actuar de forma racional: La IA esta relacionada con conductas inteligentes en artefactos (Nilsson, 1998)

¿Y LA ÉTICA?

Inclusión, diversidad y equidad implica la participación activa de una comunidad diversificada que asegure que estos sistemas se diseñan y se usan de forma tal que respetan los derechos.



<https://es.scribd.com/document/358276589/Tarea-1-Diferencias-Entre-Cada-Uno-de-Los-4-Enfoques-de-La-ia>



PROTECCIÓN DE DATOS

DR. DANIEL YANKELEVICH
DIRECTOR ORGANIZADOR
FUNDAR

PROTECCIÓN DE DATOS ¿POR QUÉ, PARA QUÉ?

El uso de datos es una herramienta imprescindible a la hora de diseñar y ejecutar políticas públicas. Hoy en día, los organismos públicos tienen varios incentivos para no compartir datos, sobre todo por los riesgos incluso penales que pueden correr si esos datos son reidentificados. Pero la solución de no compartir es la peor estrategia, porque se pierde la riqueza de la utilización de datos, que es cruzar información de diferentes fuentes: la anonimización es una metodología de trabajo recomendable porque reduce sensiblemente estos riesgos.

Vemos una tendencia a no compartir datos. Con la intención de no incurrir en problemas de violación de la privacidad, se termina perdiendo el valor y la riqueza de la utilización de datos, que es cruzar información de diferentes fuentes. No compartir datos es equivalente a apagar una computadora como medida de seguridad frente a un potencial virus informático.

Los datos son un activo para toda organización, gobierno, persona. Por lo que es importante identificar las fuentes en donde se originan los datos, explorar, contextualizar para garantizar la calidad de los datos para el conocimiento.

La Figura 1 identifica algunas fuentes de datos útiles, públicas o privadas, y lista preguntas iniciales para centrarse en el impacto y necesidad del dato (conocimiento esperado) en gobiernos, organizaciones u organismos.

Figura 1 Preguntas claves ante la oportunidad del dato: Fuentes de datos y preguntas que pueden responderse con ellas.



Generar, recolectar y almacenar datos representativos y garantizar el acceso a ellos y su utilización es fundamental para abordar problemas humanitarios y de desarrollo.

Si alguien no está representado en los datos usados para pensar y diseñar una política, es difícil que esa política lo incluya.

No se trata solo de generar datos, sino también de contar con la capacidad para utilizarlos y compartirlos. Sin embargo, existen claras limitaciones para esta práctica: no existen hoy procesos claros para compartir datos entre distintas organizaciones o incluso entre distintas áreas del sector público.

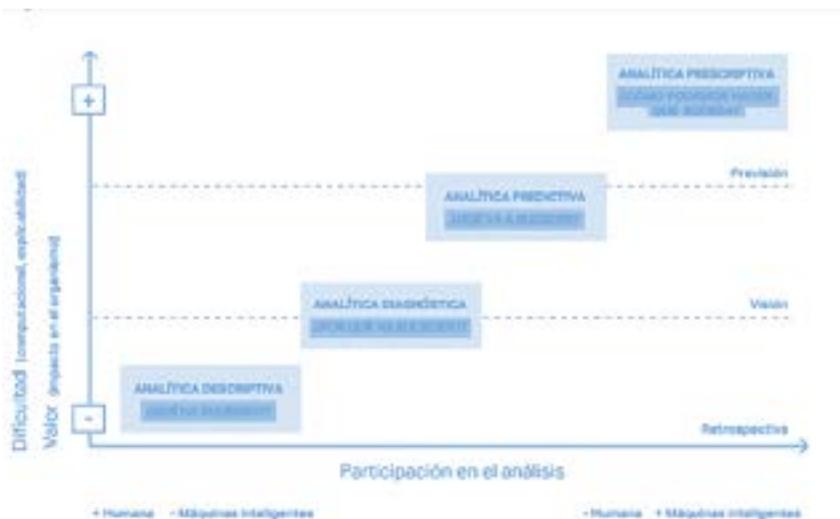
Los problemas asociados al intercambio de la información potencian la existencia de sesgos, afectan las garantías de calidad de los datos y, en consecuencia, alimentan la desconfianza en la información que los datos pueden aportar. Cuando hay falta de contexto y de transparencia, los datos y la información se vuelven fácilmente manipulables

USO DE DATOS EN LA TOMA DE DECISIONES DE SALUD DE LA NACIÓN

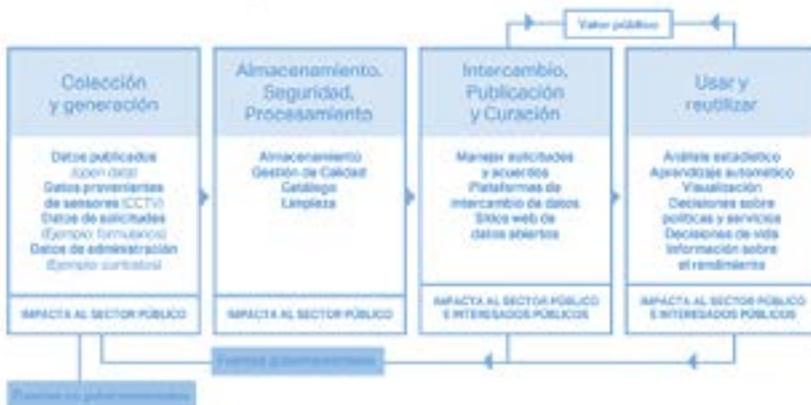
En el contexto de la pandemia de COVID-19, quedó claro que, como dijimos, el buen uso de la información salva vidas. Permitted, por ejemplo, conocer en poco tiempo dónde se producen los contagios, cortar cadenas de contagios o advertir a potenciales contactos y realizar un seguimiento cuidadoso de los pacientes.

No usar datos cuando están disponibles es negligencia.

El dato es un activo. ¿Cómo aprovecharlo? La Figura 2 muestra tipos de analítica en la era de la abundancia de datos, en base al valor del dato, complejidad de análisis y responsabilidad en la toma de decisiones. Con respecto a la participación en el análisis, en esta figura se la define por la intervención del ser humano y de máquinas inteligentes en la generación de conocimiento y responsabilidad de las consecuencias generadas.



Ciclo de valor de datos gubernamentales



Fuente: Traducción y diseño propios sobre van Dejen, C., B. Uslitt and B. Pfluy (2018), "A data-driven public sector: Enabling the strategic use of data for production, inclusive and trustful by government", OECD Working Papers on Public Governance, No. 33, OECD Publishing, Paris, <https://doi.org/10.1787/20ab162c-es>.

La sociedad debe adueñarse del mundo de datos y algoritmos y hacerlo propio para su beneficio.

IMPACTO ECONÓMICO Y SOCIAL DEL USO DE DATOS

La falta de alfabetización digital de la población en general, sobre todo respecto al uso de datos y algoritmos, conduce a una brecha o desigualdad de capacidades, lo que deja a la sociedad cautiva de quienes tienen el capital económico o humano suficiente para manejar estas tecnologías. Violencias.

MITIGACIÓN DE RIESGOS

Los impactos que el uso de datos y algoritmos pueden provocar a nivel económico y social son múltiples y revolucionarios, por eso es necesario educar a la población y a las y los tomadores de decisiones en su uso, sus oportunidades, su potencial e implicancias.



- Los datos no nacen todos iguales: hay sesgos, datos incompletos, a veces grupos enteros de poblaciones puntuales mal representados.
- La creencia de que el uso de tecnologías más avanzadas lleva directamente al desarrollo social resulta a veces bastante peligrosa, ya que no se tienen en cuenta los resguardos éticos necesarios.
- Los deepfakes, las fake news, cuyo efecto está potenciado a su vez por los algoritmos de sistemas de recomendación que nos muestran lo que probablemente seamos más

propensos a consumir, llevan a una polarización cada vez mayor de la sociedad, muchas veces a causa de información que no está basada en evidencia y que se encuentra inundada de falencias y falacias. Si a esto se suma los monopolios digitales antes mencionados, se hace cada vez más complejo encontrar fuentes de información confiable, objetiva y fehaciente



- **IDENTIFICAR DATOS:** es necesario entender qué información es confidencial o sensible y puede recuperarse en función de uno o más campos de estos datos. Por ejemplo, podría ocurrir que la identidad de la persona (su número de documento, nombre o un dato biométrico) sea el dato confidencial y se elimine de la base. Pero, como vimos en el primer ejemplo, en algunos casos puede recuperarse 10 Fundar usando fecha de nacimiento, código postal y género. De la misma forma, puede ser que no demos información en forma directa, pero que una o más variables puedan combinarse para recuperar el dato crítico con determinada probabilidad

o confianza. Finalmente, en esta etapa es necesario identificar qué propiedad de los datos se quiere conservar. Si los datos van a ser usados para construir un modelo predictivo basado en preferencias individuales, esas preferencias no pueden eliminarse ya que sería lo mismo que no compartir los datos. Entender para qué hacen falta los datos y cuál es su valor es clave para encontrar soluciones adecuadas.

- **IDENTIFICAR RIESGOS:** analizar los riesgos potenciales de la publicación, cruce y reidentificación de una base de datos es importante a la hora de priorizar y entender los escenarios de uso. El riesgo puede surgir al cruzar los datos con otras bases. El análisis no debe hacerse sólo de la base que se pretende publicar, sino de esa base en un contexto. La privacidad no es una preocupación general de la población. Varios estudios han mostrado que los jóvenes se preocupan mucho menos por la privacidad que los mayores, y que un porcentaje importante de la población está dispuesta a ceder su información a cambio de beneficios.

Los datos pueden ser considerados sensibles por diferentes razones: por revelar origen racial y étnico, opiniones políticas, convicciones religiosas, filosóficas o morales, afiliación sindical o información referente a salud o vida sexual.

- **IDENTIFICAR SOLUCIONES:** De acuerdo con el tipo de datos y los riesgos identificados en el paso anterior, se puede elegir un grado de k-anonimización que sea razonable lograr y a la vez que sea adecuado para la situación de riesgo identificada. La k-anonimización es un concepto, no una técnica.

k-anonimización: un conjunto de datos cumple con ser k-anónimo para un número k, si el mayor grado de individualización que logro usando los datos es un grupo de k individuos

Más allá de los algoritmos, que proveen un modo sistemático de tratar los datos, existen técnicas que pueden considerarse a la hora de transformar un conjunto de datos para evitar potenciales reidentificaciones:

- **Eliminación:** Siempre es posible eliminar una o más columnas o variables.
- **Generalización:** Agrupar datos, brindar información sólo a nivel de grupo (distribución estadística y parámetros generales), cambiar su nivel de detalle (granularidad). Por ejemplo, reemplazar el ingreso de un grupo familiar por el decil al que pertenece.
- **Hashing y encriptado:** Aplicar a un dato una función “de un solo sentido”, es decir, una función cuya inversa es muy costosa de calcular. Por ejemplo, reemplazar una descripción con un código, que permita identificar coincidencias pero no diferenciar entre ellas; o reemplazar un identificador con una versión encriptada, que permita identificar a los individuos pero a partir del cual no pueda recuperarse la versión original.
- **Distorsionar los datos:** (“agregar ruido”), mezclar, confundir. Esto suele ser usado en gráficos, pero se puede aplicar en varios contextos. Algunas técnicas agregan ruido aunque mantienen las propiedades relevantes de los datos originales, lo que permite aplicar técnicas de análisis de datos sin revelar información.

- **IDENTIFICAR ATAQUES O PROBLEMAS:** El factor clave en este paso es establecer un criterio adversarial. Es decir, se debe considerar que se enfrenta a un adversario y revisar la solución desde el punto de vista de un posible ataque. Además, esta visión es dinámica: aplicar una solución o una técnica seguramente genere un cambio de estrategia en el potencial ataque.

Bercovich, S.; Guaymás, A.; Penna, F. y Yankelevich, D. (2021). Datos y algoritmos para el desarrollo. Buenos Aires: Fundar. Disponible en <https://www.fund.ar>.



GOBERNANZA DE DATOS Y ADMINISTRACIÓN

LIC. CECILIA RUZ
DIRECCIÓN ANALÍTICA
DE DATOS DE AFIP

¿COMO ADMINISTRAMOS LOS DATOS?

GOBIERNO DE DATOS

- De acuerdo a la Data Management Association (DAMA, <http://www.dama.org>), data resource management (administración de datos) es el “Desarrollo y ejecución de arquitecturas, practicas y procedimientos que manejan adecuadamente las necesidades del ciclo de vida de los datos de una empresa”
- Incluye aspectos de calidad, arquitectura, seguridad y meta data de los datos.
- No es un tema de SISTEMAS es un tema de TODA la organización
- Lo datos se consideran cada vez mas un ACTIVO de la compañía

ADMINISTRAR DATOS REQUIERE TOMAR EN CUENTA VARIOS ASPECTOS:

- La seguridad
- Privacidad
- Calidad
- Semántica
- Gestión de datos maestros

CALIDAD DE DATOS

- Requiere inversión de tiempo y dinero
- Es necesario monitorearla permanentemente
- El 70% del trabajo de un proyecto es el análisis de datos

ERRORES CLÁSICOS

Fuera de rango: Edad del Paciente= 185 ()

Datos inválidos: El dato puede ser "A" o "B" pero el valor es "C"

Reglas culturales diferentes:

Fecha: Enero1, 2002 o 1-1-2002 o 1 Ene 02

Montos en diferentes monedas

Distintos formatos: (5411)674-2153 o [5411]6742153 o 54116742153

Cosméticos: Juan Lopez juan lopez

Errores de tipeo: Juan Lopez, Juna Lopez

Informar el cuit como monto de la operación

OUTLIERS

Son errores o valores atípicos

Que pasa con el monto de \$ 20.000.000 de gasto de la tarjeta de crédito?

No tirar los outliers sin examinarlos, porque pueden ser los datos más valiosos de un dataset.

Ejemplo: En 1985 tres investigadores (Farman, Gardinar y Shanklin) fueron desconcertados por un ciertos datos recopilados por el “examen antártico británico” que demostraba que los niveles del ozono para la Antártida habían caído el 10% debajo de los niveles normales de enero. El problema era, porqué el satélite Nimbo 7, que tenía instrumentos a bordo para medir con precisión los niveles del ozono, no había registrado concentraciones de ozono semejantemente bajas. Cuando examinaron los datos del satélite no les tomó mucho darse cuenta de que el satélite de hecho registraba estos niveles de concentraciones bajos y lo había estado haciendo por años. ¡Pero como las concentraciones de ozono registradas por el satélite fueron tan bajas eran tratadas como outliers por un programa de computadora y desechadas! El satélite Nimbo 7 de hecho había estado recolectando la evidencia de los niveles bajos de ozono desde 1976. El daño a la capa de ozono pasó desapercibido y no fue tratado por nueve años porque los outliers fueron desechados sin ser examinados.

DEFINICIÓN DE DATOS

En las organizaciones hay dos lugares donde típicamente se encuentran las definiciones de los datos desde el punto de vista del negocio:

- La cabeza de las personas: Estas son reglas no escritas y existen en todas las áreas de las empresas que interactúan con datos. Si las definiciones se encuentran solo en este lugar las empresas son vulnerables a la baja calidad de los datos, originada en falta de consistencia y de confianza
- En los modelos de datos: Las herramientas de modelado de datos hacen un trabajo aceptable en recolectar este tipo de información. El problema es que suelen reflejar solo el estado inicial y no los cambios.

Selecting the "Right" Meta Data to Manage, <http://www.tdan.com/view-articles/5069/>

DATOS MAESTROS

Conjunto de datos centrales que son compartidos por toda la organización. Por ejemplo: datos de clientes, de productos, de proveedores

DESAFÍO: lograr una "Golden copy" o única fuente de la verdad.

COMPARTIR INFORMACIÓN:

- Cuán confiable es la fuente de datos que me están compartiendo?
- Si tengo datos de mala calidad puedo tomar malas decisiones.
- Entendemos bien los datos? Todos entendemos lo mismo?
- Cuántos empleados tiene una empresa?
- Es importante compartir información y hacerlo dentro de un marco adecuado.

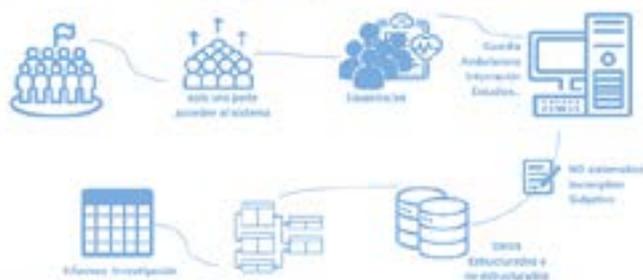
Registro Nacional de las Personas, fuente única de información de las personas



USO DE DATOS DE SALUD

DRA. VERÓNICA XHARDEZ
INVESTIGADORA DE BIG DATA

Ciclo del dato de salud (simplificado) HCE

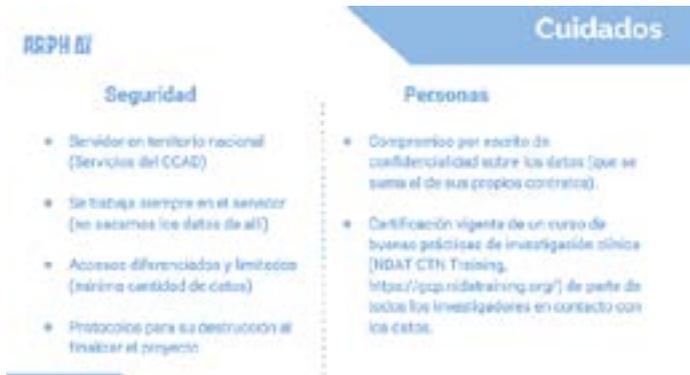
Desafíos del uso de datos en
nuestros proyectos de
Investigación

Proyecto

- Dimensión ética considerada desde la planificación.
- Aprobación de un Comité de Ética Independiente
- Línea emergente de (un Responsable de Datos (transversal a todo el proyecto)) Contribución desde la experiencia y desde los productos.

Datos

- Salen de la jurisdicción de-identificados (en datos personales)
- Suman una capa de anonimización de fuerte libre en su origen
- Los Investigadores no tienen acceso a los datos en ese procesamiento previo
- Vujan a nuestros servidores encriptados



- Los datos (en general) no están en el vacío y su uso no es neutral: detrás de los datos hay personas con derechos
- Como se recaban, como se ordenan, como se clasifican o estandarizan pueden tener implicaciones negativas.
- Los datos pueden contribuir a la toma de decisiones
- O, como ARPHAI, los datos de HCE pueden ayudar a la mejora de la gestión de epidemias, construyendo herramientas de IAyCD (sin perder autonomía).

El valor de los datos puede realizar aportes significativos en las políticas de salud. Para ello siempre deben considerarse los derechos, la buena gobernanza, la calidad, advertir sesgos y compartirlos de manera segura.



IMPACTO REGULATORIO

DELIA RODRIGO
ESPECIALISTA SENIOR EN
MEJORA REGULATORIA BID

IMPACTO REGULATORIO

¿QUE ES EL AIR? (ANÁLISIS DE IMPACTO REGULATORIO)

- Un proceso mediante el cual se analiza un problema y se identifican las opciones que existen para resolverlo, analizando los costos y beneficios de cada una de ellas, para determinar cuál es la mejor
- El AIR es un documento que ayuda al proceso de toma de decisiones y se debe construir con la participación de los grupos afectados por la problemática.
- El AIR debe realizarse antes de tomar decisión, para contribuir a la mejora del proceso regulatorio al hacer más efectivo y eficiente el diseño de las regulaciones

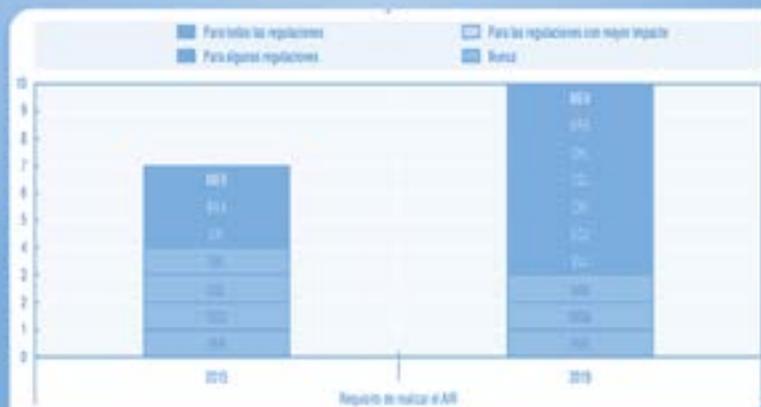
RELEVANCIA DEL AIR Y MEJORES PRÁCTICAS

- Los reguladores no tienen información perfecta, necesitan de insumos para entender el sector que regulan e identificar la mejor forma de intervenir
- El AIR ofrece datos y evidencia, ayudando a tomar mejores decisiones
- La identificación y cuantificación adecuada de costos y beneficios es importante para garantizar que la regulación propuesta trae beneficios para la sociedad
- El proceso de elaboración del AIR contribuye a consultar las partes interesadas, con lo cual se favorece la transparencia, la rendición de cuentas y un mayor cumplimiento

PASOS EN LA ELABORACIÓN DE UN AIR



USO DEL AIR EN AMÉRICA LATINA



OEI (2019). Análisis de las intervenciones públicas en América Latina y el Caribe 2010-19. O. E. I. de París.

RECOMENDACIONES PARA INTEGRAR EL USO DE AIR EN EL PROCESO LEGAL Y REGULATORIO

- Las áreas técnicas deberán preparar el AIR
- Si hay apoyo político, ese análisis será más relevante todavía porque ayudará a tomar mejores decisiones
- Hay que hacer uso de toda la información y datos que puedan obtenerse
- Hay que diseñar y planear una estrategia para vincular a los actores relevantes y abrir el proceso a la sociedad

Hay que ser consistente en la metodología, pero no complicarla demasiado

GRACIAS





OBSERVATORIO
DE FAMILIAS Y JUVENTUDES



DIPUTADOS
ARGENTINA