UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE CIENCIAS ECONÓMICAS

DOCTORADO

TESIS

GESTIÓN DE LA PRIVACIDAD DE DATOS EN ORGANIZACIONES PÚBLICAS DEL ESTADO ARGENTINO. ELEMENTOS ESTRUCTURALES CLAVES PARA EL DISEÑO DE UNA ESTRATEGIA RESPONSABLE.

ALUMNA: NATALIA R. SALABERRY

DIRECTOR DE TESIS: DR. JAVIER I. GARCÍA FRONTI

CODIRECTORA DE TESIS: DRA. MARÍA J. BIANCO

MIEMBROS DEL TRIBUNAL DE TESIS: MARÍA TERESA CASPARRI, RUBEN FUSARIO, GASTÓN MILANESI

FECHA DE DEFENSA DE TESIS: 19 DE JUNIO DE 2025

Agradecimientos

Quisiera expresar mi profundo agradecimiento a mi director de tesis, Dr. Javier García Fronti, y a mi subdirectora, la Dr. María José Bianco, por el acompañamiento, confianza y apoyo incondicional para llevar adelante este trabajo. Sus continuos aportes realizados fueron fundamentales para la elaboración de esta tesis. Sin ellos, no hubiese sido posible.

También quisiera mencionar mi enorme gratitud hacia la Profesora Emérita Dra. María Teresa Casparri. Su trabajo incansable, su apoyo y motivación constante, fueron claves fundamentales para culminar este trabajo.

Agradezco a la Universidad de Buenos Aires, a la Facultad de Ciencias Económicas, por permitirme ser parte de un proceso continuo de formación. El conocimiento brindado a través de todos sus docentes en los diferentes niveles de formación académica son el eje fundamental para la creación de inclusión y sentido de pertenencia. Gracias a ellos y ellas puedo decir que esta Facultad siempre será mi hogar.

Al Centro de Métodos Cuantitativos Aplicados a la Economía y a la Gestión perteneciente al IADCOM, por interesarse en mi propuesta y permitirme continuar con el desarrollo de mi trabajo como investigadora académica en formación.

Finalmente, agradecer a mi familia, pilar fundamental en mi vida. En especial a mi madre, que con sus valores y amor incondicional siempre buscó potenciarme en el camino que he elegido seguir. Pero sobre todo por haberme enseñado que, con perseverancia, compromiso y trabajo a pesar de las adversidades, los objetivos se pueden cumplir.

Resumen

Un pilar fundamental de la gestión de la privacidad de datos en contextos organizacionales es la responsabilidad en la gobernanza del riesgo asociado al uso de datos personales. Mediante la utilización de tecnología las organizaciones convierten a los datos generados por los individuos en información que es utilizada para la toma de decisiones. La valorización producida se configura en un proceso complejo dando lugar a un nuevo mercado de datos en un entorno digital. Si bien en él se produce un intercambio ágil, también se genera una asimetría de poder que expone la vulnerabilidad del derecho a privacidad de los ciudadanos.

El impacto causado en la forma de hacer políticas públicas por este nuevo escenario presenta beneficios, pero también reta a las organizaciones estatales a asumir nuevas responsabilidades. Frente al potencial de la tecnología para la generación de conocimiento en base a datos personales, resulta necesario que las organizaciones adopten un accionar preventivo frente a los riesgos asociados. Mediante la detección de los elementos claves para la construcción de privacidad de datos se podría alcanzar un diseño de una estrategia de gestión eficiente y eficaz bajo un marco de responsabilidad.

El objetivo general que propone esta tesis es determinar los elementos estructurales claves para el diseño de una estrategia de gestión responsable de la privacidad de datos personales en organizaciones estatales. En el primer capítulo, se analiza como la interacción entre individuos, tecnología y organizaciones conforman un nuevo mercado de datos del cual se deriva un riesgo de privacidad. En el segundo, se elabora una definición de la privacidad de datos personales que incorpora a la responsabilidad organizacional como factor clave para la construcción de privacidad desde el diseño y por defecto. En el tercero, se expone el uso de una metodología cuantitativa como herramienta para impedir la identificación de individuos en una base a datos que permite brindar cierta garantía de privacidad. Finalmente, en el cuarto capítulo se identifican los factores estructurales claves y un modelo como propuesta para su vinculación. Estos elementos permitirán mostrar el potencial para diseñar una estrategia de gestión responsable de la privacidad de datos personales.

En la elaboración del trabajo se utiliza un enfoque cualitativo y cuantitativo. A partir del primero se indaga en la discusión literaria sobre la privacidad de datos. Como resultado se induce una hipótesis general a constatar y se establece a la responsabilidad organizacional como elemento clave principal en la gestión de la privacidad de datos. Con el segundo, se

expone una metodología para brindar garantía de protección de datos personales. De la conjunción de ambos, es posible determinar los elementos claves para el diseño de una estrategia responsable de gestión de la privacidad de datos en contextos organizacionales.

La tesis aporta los elementos estructurales claves para la protección de datos personales y una propuesta para su modelización bajo un marco de responsabilidad organizacional, constituyéndose en una herramienta de gestión. La identificación de riesgos en el procesamiento de datos personales y la implementación de técnicas cuantitativas para mitigarlos permite anticipar los principales resultados potenciales de este tipo de desarrollo. Además, facilita mostrar la importancia de elaborar una estrategia de gestión responsable de privacidad de datos como aporte a la construcción del bienestar de la sociedad.

Palabras claves: Gestión de la privacidad de datos, Organizaciones públicas del Estado argentino, Elementos estructurales claves, Estrategia responsable.

Clasificación JEL: O320, C6, L8, L860

Índice

Introducción

Desde 1948, la privacidad es un derecho humano universal según lo establece la Organización de las Naciones Unidas (ONU). Con el advenimiento de la Sociedad de la Información desde fines del siglo XX, surge un nuevo espacio digital de interacción entre individuos, tecnología y organizaciones, donde emergen nuevos riesgos asociados al procesamiento de datos vinculados a la privacidad. A partir de entonces se amplía la concepción original de la privacidad para abarcar también a la privacidad de datos personales. La identificación de individuos derivada del procesamiento de datos, se la considera una violación al derecho de privacidad en todos los países miembros de la ONU.

La Sociedad de la Información se caracteriza por el accionar de los agentes a través de una red digital. El poder obtenido por las organizaciones es producido en base al uso intensivo de tecnología para convertir un bien intangible (datos) en información. Si bien esto ha beneficiado a las organizaciones públicas para la creación de valor público, también ha generado una situación de vulnerabilidad para los ciudadanos (Bell, 1979; Mattelart, 2001; Castells, 2002; Latour, 2005; Varían 2010; OCDE, 2014; Mortier et. al, 2014; Bryson et al., 2015; Moazed y Johnson, 2016; Täuscher y Laudien, 2018; Zysman y Kenney, 2018; Culpepper y Thelen, 2019; Zudoff, 2019; Mazzucato, Entsminger, & Kattel, 2020; Ghandy Jr., 2021; Naser, 2021).

La nueva configuración de un espacio digital de interacción ha dado origen a un nuevo mercado de datos. Definido como un proceso de co-creación a partir de la interacción digital entre humanos y no humanos (Latour, 2005; García Fronti y Herrera, 2020), se produce una asimetría de información causada por la existencia de un proceso continuo y omnipresente de datos. Este proceso es creado y controlado por las organizaciones, siendo el sustento de su poder de acción superior. La consecuencia directa es que en ocasiones es utilizado como medio para modificar el escenario de acciones posibles de los individuos. Además, la recopilación de los datos personales suele ser sin el consentimiento de su titular. A partir de ello surgen diferentes posturas entorno a la privacidad asociada al uso de datos personales (Smith, Dinev, & Xu, 2011; Friedman, Kahn, Borning & Huldtgren, 2013; Mortier et.al., 2016, Zudoff, 2019; Ghandy Jr., 2021) sentando las bases de diferente regulación.

En un intento de comenzar a controlar la problemática en torno a la privacidad de datos personales en el nuevo contexto digital, ha surgido diferente regulación. Entre estas, se

encuentran dos que presentan enfoques muy diferentes. Por un lado, la norteamericana que posiciona al individuo como un consumidor y utiliza la herramienta legislativa como soporte para asignarle la responsabilidad del control de su privacidad. Por el otro, la regulación europea que reconoce a la privacidad como un derecho humano y busca contrarrestar la concentración en manos privadas (Varían & Berkeley, 1996; Smith, Dinev, & Xu, 2011; Rubinstein, 2012; RGPD, 2016; Arner, Castellano y Selga, 2022). Entre medio de estas, se encuentra la legislación argentina con su Ley 25.326 de Protección de Datos Personales y su marco constitucional.

La ley 25.326 fue sancionada en el año 2000 a partir de la cual el uso de datos personales en contextos organizacionales quede efectivamente regulada o normada. Si bien constituye una herramienta legislativa que brinda cierta garantía legal a los ciudadanos, en ningún momento menciona el término privacidad de datos personales. Por un lado, esto marca la desactualización que posee ante el avance de las tecnologías de la información y los cambios sociales que ello ha producido durante las dos décadas posteriores a su promulgación. Por el otro, al ser un instrumento que se hace efectivo luego de que un hecho (o daño) ocurrió, es posible interpretar que la responsabilidad del control de los datos recae únicamente sobre el individuo.

Dada la mencionada legislación argentina, pareciera no ser garantía suficiente para alcanzar privacidad en base a la protección de datos personales. Diferentes hechos de conocimiento público como ser el robo de datos personales de jubilados de la base de datos de PAMI (Instituto Nacional de Servicios Sociales para Jubilados y Pensionados) en agosto de 2023[1], la violación de acceso a servidores del poder judicial de la provincia de Chaco en enero de 2022[2] o la denuncia presentada por el uso irrestricto de datos biométricos en un organismo del Gobierno de la Ciudad de Buenos Aires en 2023[3], entre otros, exponen el riesgo al cual quedan expuestos los ciudadanos cuando sus datos son vulnerados. En este sentido, la ley no provee un enfoque de prevención sino más bien actúa reactivamente.

El caracter reactivo de una norma o ley, implica un accionar posterior a que un incidente haya expuesto el problema. Si bien, brinda ciertas garantías para una convivencia en

[1] https://www.cronista.com/economia-politica/jubilados-pami-se-filtraron-los-datos-de-los-beneficiarios-que-cuidados-tomar/
[2] https://www.iproup.com/innovacion/28826-hackean-los-servidores-del-poder-judicial-de-chaco
[3] https://www.ambito.com/politica/denuncian-penalmente-la-ciudad-el-uso-datos-biometricos-justificacion-racional-n5705668

sociedad, pareciera no ser suficiente para resolver la problemática de privacidad de datos. A partir de ello se abre la posibilidad de plantear él porque es necesario una gestión responsable de datos personales en organizaciones. Desde una perspectiva de responsabilidad en torno al respeto del derecho humano y de uso de tecnología sobre datos personales, esto podrá ser sintetizado en la idea de privacidad desde el diseño y por defecto. El elemento clave de esta concepción es la responsabilidad involucrada en las organizaciones a la hora de diseñar procesos para la recopilación y procesamiento de datos personales. Pero también, surge la necesidad de implementar métodos cuantitativos. Por un lado, para brindar garantías de protección de los datos. En este sentido, entre los métodos existente, la Privacidad Diferencial ofrece una propuesta que puede contribuir a lograrlo (Dwork, 2008; Cavoukian, 2011; AEPD, 2019, 2020). Por el otro, para modelizar la contribución de la responsabilidad a la construcción de privacidad.

En este contexto, surge el siguiente interrogante: ¿Cómo pueden protegerse los datos personales recopilados y publicados por las organizaciones del Estado argentino garantizando la privacidad de los individuos? Para poder responder este interrogante, el objetivo general de la tesis es determinar los elementos estructurales claves para el diseño de una estrategia de gestión responsable de la privacidad de datos personales en organizaciones estatales argentinas. La hipótesis asociada a este objetivo es que el procesamiento y uso de datos personales por parte de las organizaciones estatales argentinas para la elaboración de políticas públicas vulnera la privacidad de los individuos.

Para llevar adelante el desarrollo del objetivo general propuesto en esta tesis, se plantean cuatro objetivos específicos. El primero es analizar la interacción entre organizaciones, individuos y tecnología en el mercado de datos y como el uso de datos personales por parte de las primeras puede vulnerar la privacidad de los ciudadanos. Asociado a lo anterior, surge el segundo objetivo específico que consiste en elaborar una definición de privacidad de datos personales que incorpore a la responsabilidad organizacional como clave para mitigar el riesgo de privacidad. Para poder abordar en una primera instancia la construcción de responsabilidad organizacional, el tercer objetivo específico que se plantea es establecer a la Privacidad Diferencial como un método cuantitativo que brinda cierta garantía de protección de datos personales. Finalmente, considerando a los anteriores, el cuarto objetivo específico consiste en especificar los factores estructurales claves para el diseño de una estrategia de gestión responsable de la privacidad de datos personales en organizaciones públicas

argentinas. Con el fin de poder cumplimentar con cada uno de ellos, la realización de la tesis se organiza en cuatro capítulos, agrupados en dos partes.

La primera parte que comprende los capítulos 1 y 2, se expone el corpus teórico de la tesis. Mientras en el primero se indaga en la conformación del mercado de datos y el riesgo derivado para la privacidad de los individuos; en el segundo se incorpora a la responsabilidad organizacional para elaborar una definición de la privacidad de datos personales. Se explica entonces, como la utilización de tecnología para el procesamiento de datos, si bien tiene beneficios para la elaboración de políticas públicas eficientes y eficaces, también requiere de incorporar a la responsabilidad organizacional para poder brindar garantías completas de protección de datos personales.

El objetivo del primer capítulo es analizar cómo se conforma un nuevo mercado de datos a partir de la interacción entre individuos, tecnología y organizaciones que puede vulnerar la privacidad de los primeros. La hipótesis asociada a este capítulo es que el procesamiento de datos personales por parte de las organizaciones públicas estatales pone en riesgo la privacidad de los individuos. Al explicar al mercado como un proceso de co-constitución entre los actores intervinientes, surge un riesgo de violación de la privacidad en base a la utilización de datos personales por parte de las organizaciones.

Para poder argumentar porqué la responsabilidad organizacional emerge como relevante en este escenario, se coloca al ciudadano como parte central de los sistemas de datos. Bajo esta perspectiva, la capacidad de cálculo de organizaciones y ciudadanos son asimétricas. Ello porque las primeras cuentan con una capacidad predictiva y de gestión de datos superior a la del individuo ya que posee control y manejo de tecnologías de *Big Data*. Si bien, tal capacidad de acción facilita la generación de conocimiento para la elaboración de políticas públicas, también puede ser utilizada para modificar el escenario de acciones posibles del individuo. La asimetría de información producida da origen y sustento a una asimetría de poder de control por parte de las organizaciones públicas del Estado que puede ser cuestionado en términos de privacidad.

El objetivo del segundo capítulo es elaborar una definición de la privacidad de datos personales que incorpore a la responsabilidad organizacional como factor clave para la construcción de privacidad desde el diseño y por defecto. La hipótesis planteada en este capítulo es que la responsabilidad organizacional para abordar la problemática de la

privacidad de datos personales en términos operativos es necesaria para mitigar el riesgo asociado. En el capítulo se complementa y consolida la argumentación planteada en el capítulo 1 incorporando la perspectiva regulatoria que considera a la privacidad como un derecho humano. En primer lugar, se amplía la concepción de la privacidad física hacia la privacidad asociada a datos personales. Esto permite abordarla desde un enfoque normativo y operativo, donde la responsabilidad organizacional juega un rol clave.

Abordar una definición de la privacidad de datos personales desde el diseño y por defecto, requiere incorporar a la responsabilidad operativa organizacional. Partiendo desde una concepción normativa de la privacidad en tanto derecho humano universal y, dado el procesamiento omnipresente de datos en el mercado de datos, se argumenta una definición de la privacidad de datos personales. De aquí se desprende que, para alcanzar una eficacia operativa en la mitigación del riesgo de privacidad asociado al uso de datos personales en un contexto organizacional, las posturas dominantes de regulación como la norteamericana y la europea, y la existente en Argentina, resultan necesarias pero insuficientes. Frente a ello se realiza un abordaje desde un enfoque de gestión de datos basada en principios operativos. En este sentido, se propone como acción base la utilización de una metodología cuantitativa para brindar garantía de protección de datos personales a la hora de diseñar los procesos de datos en las organizaciones.

A partir de la vinculación entre las dos hipótesis abordadas en el corpus teórico de la tesis, se definió la hipótesis general. Se enuncia como sigue. El procesamiento y uso de datos personales por parte de las organizaciones estatales argentinas para la elaboración de políticas públicas vulnera la privacidad de los individuos. Si bien, las organizaciones cumplen con cierta regulación establecida, no resulta suficiente para brindar garantías efectivas de privacidad a los ciudadanos en el contexto del mercado de datos. Ante este escenario, la responsabilidad operativa en el tratamiento y uso de datos personales por parte de estas resulta clave. La aplicación de un método cuantitativo puede contribuir en este sentido.

En la segunda parte de esta tesis se desarrollan los capítulos 3 y 4. En el primero se presenta, evalúa y establece un método cuantitativo para la protección de datos personales. En el capítulo 4 se determinan los elementos estructurales claves para la construcción de una gestión responsable de datos personales en organizaciones. Mientras en el capítulo 3 se utiliza un enfoque cuantitativo para exponer la efectividad de la metodología de Privacidad

Diferencial en términos de evitar la identificación de individuos en una base a datos personales; en el último capítulo se utiliza un enfoque cualitativo para la determinación de los factores principales para el diseño de una estrategia responsable de privacidad en organizaciones y un enfoque cuantitativo para modelizar su vinculación.

El objetivo del capítulo 3, es presentar una metodología cuantitativa para impedir la identificación de individuos en base a datos sensibles que otorga garantía de privacidad a los ciudadanos. La hipótesis establecida en este capítulo es que el método de la Privacidad Diferencial es efectivo para la protección de datos personales. En el capítulo, primero, se aborda su concepto. Se trata de un método matemático probabilístico que permite aplicar ruido aleatorio a los datos que brinda cierta garantía de privacidad. Pero también resulta importante que permita encontrar un equilibrio entre la necesidad de uso de los datos y su protección. Así podrá ser considerada para contribuir a llevar adelante una gestión responsable.

Para poder ejemplificar la operatoria del método de Privacidad Diferencial, en primer lugar, se presenta su modelización desde sus fundamentos teóricos y los principales conceptos involucrados para su comprensión. Seguidamente se aplica sobre un conjunto de datos personales. La base de datos se construye manualmente a partir de tomar una muestra de un listado de CUIT (Clave Única de Identificación Tributaria). Con cada uno estos se consultan diferentes servicios *web* de organizaciones públicas del Estado argentino para la obtención de otros atributos personales. Finalmente, mediante la utilización del lenguaje de programación *Python*, se lleva a cabo su aplicación desarrollando un algoritmo y posteriormente se realiza una evaluación de resultados. Como resultado de la aplicación, se concluye que la metodología tiene un potencial adecuado para evitar identificar individuos en una base de datos. A su vez, su aplicación permite identificar parámetros claves para la construcción de privacidad.

El objetivo del capítulo 4, es determinar los elementos estructurales claves para llevar adelante una gestión responsable de datos personales en organizaciones. Si bien el capítulo no posee una hipótesis especifica, se deriva como consecuencia de los desarrollado previamente. En este sentido se utiliza un enfoque cualitativo para la determinación de los elementos claves y relevantes que surgen de los resultados previos alcanzados. La responsabilidad organizacional y los parámetros del modelo de Privacidad Diferencial son los principales elementos que emergen como relevantes para el diseño de una estrategia

responsable de datos personales. Finalmente, la vinculación entre responsabilidad y privacidad de datos personales es alcanzada mediante una modelización propuesta. En base a su aplicación sobre una base de datos, se evalúa como la responsabilidad contribuye significativamente en la construcción de privacidad de datos. Con los resultados preliminares obtenidos, puede interpretarse que la responsabilidad es clave para la construcción de un marco operativo de gestión responsable.

Tras el cumplimiento total y parcial de cada uno de los objetivos enunciados, la tesis aporta un conjunto de elementos claves para complementar la normativa vigente en la protección de datos personales en organizaciones. Esto se constituye en una herramienta fundamental para llevar adelante el diseño de una gestión responsable de datos. La determinación de la responsabilidad involucrada por parte de las organizaciones en la protección de datos como elemento esencial para brindar garantías de privacidad a los ciudadanos, permite anticipar el potencial de los resultados alcanzados. En este sentido, se espera por un lado que las organizaciones, particularmente las públicas del Estado argentino, puedan acceder a implementar lo propuesto. Además, ante la posibilidad de una revisión para la actualización de la normativa vigente, pueda ser contemplado.

Capítulo 1: La sociedad de la información, las organizaciones y el mercado de datos

Introducción

La disrupción de las tecnologías desde finales del siglo XX ha dado origen a lo que se denomina Sociedad de la Información. Caracterizada por el accionar de los agentes a través de una red digital (Bell, 1979; Mattelart, 2001; Castells, 2002; Latour, 2005; Mortier et. al., 2014), comienza a visibilizarse como el poder es producido en base al uso intensivo de tecnología para convertir un bien intangible (datos) en información. Si bien esto ha beneficiado a las organizaciones públicas para la creación de valor público (Bryson et al., 2015; Naser, 2021), se ha generado una situación de vulnerabilidad social para los ciudadanos (Mortier et. al., 2014; Zudoff, 2019; Ghandy Jr., 2021).

Una de las características de Sociedad de la Información, es la creación de un nuevo entorno de interrelaciones que recibe el nombre de plataformas digitales (Varían 2010; Moazed y Johnson, 2016; Zysman y Kenney, 2018; Culpepper y Thelen, 2019). Estas son un espacio digital donde se produce la interacción entre individuos y organizaciones con diferentes fines. En el contexto estatal, a tal plataforma se la conoce como Gobierno Digital (OCDE, 2014). En este, la capacidad de generar valor en base a datos de las organizaciones públicas se ha visto potenciada por su rápido crecimiento y masiva disponibilidad. Pero al mismo tiempo ha limitado la capacidad de acción de los individuos (Zuboff, 2019, Ghandy Jr., 2021).

La nueva configuración del espacio digital da forma a un nuevo mercado: el de datos (Täuscher y Laudien, 2018; Mazzucato, Entsminger, & Kattel, 2020). Definido como un proceso de co-creación a partir de la interacción digital entre humanos y no humanos, la creación de valor público también es posible (Bryson et al., 2015). A su vez, al tratarse de un mercado se produce una asimetría. En este mercado en particular, se trata de una asimetría de información que se origina con la creación de un proceso continuo a la vez que omnipresente de datos por parte de las organizaciones. Como consecuencia, las organizaciones públicas se posicionan con un poder superior frente al ciudadano que toma forma de vigilancia o de control que limita su capacidad de acción (Zuboff, 2019; Mazzucato, Entsminger y Kattel, 2020; Ghandy Jr., 2021).

En este contexto, surge el siguiente interrogante: ¿cómo se construye el poder de acción superior en base a datos en las organizaciones públicas del Estado y cuáles son las

implicancias para los individuos? Para responder esta pregunta el objetivo a desarrollar en el presente capítulo consiste en analizar la interacción entre organizaciones estatales e individuos en el mercado de datos. Para ello resulta imprescindible colocar al individuo en el centro de la sociedad de la información y comprender el proceso constitutivo del mercado, como se crea valor público y como se produce una asimetría de poder en favor de las organizaciones.

A fin de realizar una comprensión integral del Estado como una organización basada en datos y las consecuencias derivadas de su interacción con los ciudadanos, el presente capítulo se divide en tres secciones. En la primera sección se desarrolla la conceptualización de la sociedad de la información para comprender el contexto social en el que se desarrolla el gobierno digital y sus implicancias. Esta es abordada a partir de tres dimensiones claves: la sociedad en red, la interacción de seres humanos con datos y el poder producido en base a información.

La segunda sección, se centra en la concepción de una organización pública que basa sus decisiones en datos a través del uso de tecnología. En primer lugar, se desarrolla su definición a partir del poder de cálculo predictivo en base al uso intensivo de tecnología de *Big Data*. Luego, se específica sobre la importancia y necesidad de una gestión de datos. Finalmente se analiza cómo se lleva a cabo la generación de conocimiento para la toma de decisiones en base a información mediante la utilización de tal tecnología.

En la última sección, se analizan las principales particularidades del mercado de datos que se genera en el contexto de la sociedad de la información. En primer lugar, se aborda la conformación de este mercado como un ensamblaje sociotécnico dentro de un proceso de co-constitución. Se continúa con el desarrollo sobre la creación de valor de mercado, en particular de valor público. Por último, se especifica sobre la asimetría de información producida que da origen y sustento a una asimetría de poder de vigilancia o control por parte de las organizaciones públicas del Estado.

1.1 Caracterización de la Sociedad de la Información

Con la evolución tecnológica de finales del siglo XX se produce un cambio en la configuración social que da origen a lo que se conoce como Sociedad de la Información. Se la define como aquella caracterizada por el accionar de actores en red mediante el uso de tecnología para la generación de información. En esta, el poder producido genera una

desigualdad en el ejercicio de la capacidad de acción del individuo (Bell, 1979; Mattelart, 2001; Castells, 2002; Latour, 2005; Mortier et. al., 2014).

En esta concepción de la Sociedad de la Información se destacan inicialmente tres elementos que son la tecnología, la información y la capacidad de acción. Mientras la primera se conforma por el conjunto de recursos físicos e intangibles (*Hardware* y *Software*, procesos, conocimiento, datos) (Haskel y Westlake, 2017), la segunda es el resultado obtenido de la aplicación de conocimiento a través de la primera (Bell, 1979; Castells, 2002). Esto solo es posible gracias a la capacidad de acción de los actores intervinientes en la red de interacción (Latour, 2005).

Los actores se concentran en dos figuras principales, las organizaciones y los individuos. Las primeras, concentran los recursos, mientras que los segundos son los ciudadanos generadores de datos. Ambas poseen capacidad de acción, pero el poder de acción obtenido por las organizaciones en base a la generación de información limita el accionar del individuo colocándolo en una situación cuanto menos de desventaja (Mortier et. al., 2014, Zudoff 2015, 2019; Gandy Jr, 2021).

La concentración de poder en el actor que posee la capacidad de modificar el comportamiento humano o restringir las opciones de acción del ciudadano (Zuboff, 2015; Gandy Jr, 2021) invita a pensar que es sostenido por un proceso de generación de información omnipresente en esta nueva configuración social en red. Con el fin de desarrollar esta idea, a continuación, se comienza por abordar el concepto de sociedad de la información.

1.1.1 La sociedad de la información como una red

En la literatura, es frecuente encontrar el término Sociedad de la Información para referirse, principalmente, al cambio social de finales del siglo XX y comienzo del siglo XXI. Podría decirse que su característica distintiva viene dada por la disrupción de las Tecnologías de la Información y Comunicaciones (TIC)[4]. Su uso generalizado, ha impactado no solo en la forma en que los individuos se relacionan, sino también en la manera en que la información es generada.

[4] La UNESCO define a las TIC como "Conjunto diverso de herramientas y recursos tecnológicos utilizados para transmitir, almacenar, crear, compartir o intercambiar información"

El concepto “Sociedad de la Información” fue empleado por primera vez por el sociólogo norteamericano Daniel Bell (1919-2011) en su trabajo “*The Social Framework of the Information Society*” de 1979. En este, Bell la caracteriza a partir de tres dimensiones. La primera de ella se refiere al paso de una sociedad productora de bienes a una sociedad de servicios. Particularmente, observando el caso de la sociedad norteamericana, destaca, por un lado, a los servicios brindados por profesionales basados en la programación y procesamiento de información mediados por computadoras. Por el otro, a los servicios sociales como ser la enseñanza, la salud, entre otros. Pero el eje central de su conceptualización es la segunda dimensión propuesta por el autor.

La segunda dimensión propuesta por Bell implica la codificación del conocimiento teórico como conductor del cambio social. Es decir, es la fusión entre ciencia y tecnología que hace a la modificación de la propia tecnología (Bell, 1979). La tecnología ya no es utilizada únicamente para la producción de bienes como en la sociedad industrial. Ahora, empleada en conjunto con la ciencia, facilita la generación de nueva tecnología y conocimiento. En este sentido, es que se establece una relación funcional de la tecnología que brinda la posibilidad de obtener un poder explicativo.

En este contexto, el poder explicativo a través de la utilización de tecnología se encuentra relacionado con la capacidad de codificar la información para estructurar el conocimiento. La posibilidad de procesar grandes cantidades de información para transformarla en dígitos numéricos permite aplicar métodos cuantitativos derivados principalmente de la matemática. A partir de ello es posible realizar cálculos inicialmente exploratorios. Esta manera estructurada de explorar la información facilita la obtención de un mayor conocimiento sobre un problema dado y se podrán tomar decisiones más fundamentadas.

Quienes llevan adelante la tarea de decidir en una organización, ahora pueden contar con un mayor conocimiento del problema a resolver o del contexto en el cual se produce cierto fenómeno. Las decisiones intuitivas son sustituidas por decisiones informadas en base a los resultados obtenidos de un proceso algorítmico que facilita la racionalidad de la acción (Mattelart, 2001). Los algoritmos pueden ser definidos como una secuencia de pasos lógicos dados a una computadora como instrucciones a través de los cuales se obtiene un resultado que es materializado en la toma de decisiones. De esta manera, se configura una relación entre personas y ciencia de carácter cooperativo y recíproco.

La tercera dimensión descripta por Bell (1979) refiere a que el nuevo poder explicativo se convierte en el elemento clave para la toma de decisiones en organizaciones. De este modo, su propuesta se centra en una "tecnología intelectual" que se convierte en la forma predominante de gestión en las organizaciones (Bell, 1979; Mattelart, 2001). Así, la sociedad de la información es una sociedad funcional que está gestionada por los principios científicos (Mattelart, 2001) a través de tecnología. Esta última viene dada por los algoritmos que poseen mayor poder predictivo. Podría decirse entonces que se abre el camino a una justificación tecnocientífica de la sociedad, cuya clave es la combinación de tecnología y especialistas que poseen conocimiento científico en un contexto organizacional.

Unos años más tardes, en 2002, tomando la definición tecnológica de Bell, Manuel Castells[5] escribiría que en realidad el cambio producido en la sociedad del siglo XXI, no se centra en la tecnología y el conocimiento, sino en "la aplicación del conocimiento e información a aparatos de generación de conocimiento y procesamiento de la información/comunicación" (p.61, Vol. I). En esta concepción, es posible encontrar la referencia a un proceso de generación de información para la obtención de conocimiento, ampliando el concepto de sociedad de la información. Ya no solo se trata de aplicar tecnología y conocimiento sino de desarrollar procesos para la generación de conocimiento. Y estos procesos es posible llevarlos a cabo porque se actúa en red.

De este modo, Castells propone concebir al cambio social de fines del siglo XX como una red. La configuración de los procesos de generación de conocimiento viene dada por la interconexión de nodos, siendo posible gracias a las tecnologías de la información (Castells, 2002). Los nodos serán específicos de cada red, y luego existirá un punto de intersección. Concretamente, cuando confluyen la tecnología, el conocimiento científico y los recursos de capital necesarios en la red de la organización, entonces es posible obtener como resultado la generación de nueva información o conocimiento.

Un interrogante que surge es cómo se distribuye el nuevo conocimiento generado en esta configuración social. Una respuesta posible se puede encontrar en el rol que cumple Internet. Se define como un mecanismo para la difusión de información y un medio para la colaboración e interacción entre las personas y sus computadoras sin importar su ubicación geográfica (Leiner, B. et al., 2009). Pero Castells (2002), agrega una dimensión cultural para

[5] sociólogo español, nacido el 9 de febrero de 1942 en Hellín, municipio de una ciudad española de la provincia de Albacete, en la comunidad autónoma de Castilla-La Mancha, situada al sureste de la península ibérica.

su concepción. Si bien Internet requiere de tecnología informática para su funcionamiento, sostiene que en su desarrollo existió la necesidad de pensar la idea que le dio origen. Hubo una necesidad "social" de desarrollar el instrumento de comunicación necesario para que la red del conocimiento se interconecte (Castells, 2002).

Bajo esta concepción de red es posible encontrar a la tecnología y a la información (conocimiento) como características esenciales de su construcción. Si bien, en la actualidad su rol es clave y de suma relevancia, se considera que no tiene en cuenta el resultado que surge de la acción de los actores que intervienen. En este punto resulta interesante la propuesta del filósofo francés Bruno Latour. Prácticamente una década y media posterior a Bell y tres años posterior a Castells, Latour (2005) escribiría sobre la necesidad de pensar a la red como múltiples asociaciones entre actores y entidades que van dejando rastros en su accionar.

Desde la perspectiva de Latour, podría decirse que la red es un concepto y no una cosa que se encuentra allí afuera que va de lo individual a lo social colectivo. Para facilitar su comprensión, Latour (2011) nos brinda un claro ejemplo. Los servidores creados por los ingenieros para que el buscador *Google* funcione permiten comprender cómo el rastro (acción) se materializa y se hace visible para entender porque funciona. *Google* no existiría sin servidores funcionando. Y estos, a su vez, no existirían si los ingenieros no los hubiesen pensado y desarrollado a partir de su conocimiento especializado y los recursos disponibles. Con ello se hace visible el objeto, el servidor, y de ahí la posibilidad para la mente humana de comprender su funcionamiento. Así, el elemento clave del funcionamiento de la red es la capacidad de acción de los actores humanos y no humanos.

De este modo, es posible concebir a la red a partir del conjunto de atributos (los rastros) que los actores van dejando en su accionar dentro de la misma. A partir de ello, la sociedad de la información puede ser comprendida como una red que viene configurada por el accionar de los actores visible a través de los rastros que dejan al interactuar entre sí. Particularmente, en un contexto organizacional, esto es posible de observarse cuando se lleva a cabo el proceso de aplicación de conocimiento especializado sobre datos a través de tecnología obteniéndose como resultado la generación de nuevo conocimiento. Pero a su vez, el conocimiento generado es colaborativo y distribuido a través de un medio digital como internet. Este permite que la red se retroalimente continuamente ampliando cada vez más la capacidad de acción de los actores.

Ahora bien, la ampliación de la capacidad de acción implica la obtención de un poder superior por parte de la organización. Este poder es producido (Latour, 2005) siendo que ya no se deriva de la clásica acumulación de capital sino de dicha capacidad. Pero a su vez, los individuos o ciudadanos que interactúan con la organización ven su poder de decisión disminuido quedando en una situación de desventaja. Esto debido a que no son parte del sistema interno de datos de la organización, sino solo quienes generan los datos posteriormente apropiados. En el siguiente apartado se profundiza sobre este punto.

1.1.2 Interacción entre actores y datos

En la red definida en el apartado anterior, el recorrido que realizan los datos (rastros) es un proceso complejo a la vez que frecuente. A su vez, el mecanismo para su obtención es omnipresente. El dato es el valor que toma una variable bajo análisis y que son generados por la acción humana de forma individual o colectiva en su accionar (Gil Flores et. al., 1994). En este sentido, se trata de una unidad mínima de información que se encuentra dispersa en la red. Se requiere de recolectarlos y procesarlos para generar información.

Entre los actores intervinientes en la red de la sociedad de la información, se encuentran la tecnología, individuos especializados y los ciudadanos. Los especialistas interactúan con datos para su procesamiento. Los ciudadanos, son quienes generan datos. En ambos casos, el accionar de los actores es a través de la utilización de tecnología. Mientras en el primer caso es para la construcción de información; en el caso de los ciudadanos en general es para la obtención de un servicio. Bajo este esquema de interacción, resulta insuficiente considerar únicamente como interactúan con los dispositivos tecnológicos con el fin de alcanzar una comprensión más acabada. Si no, también, resulta necesario preguntarse cuáles son las consecuencias que surgen.

En este sentido, Mortier et. al. (2014), propone colocar al ciudadano generador de datos en el centro de su circulación para considerar el impacto causado. Para ello, identifica tres puntos clave entorno al uso de datos: la transparencia, la capacidad de acción y la alineación de normas sociales. Cada uno de estos aspectos permiten ver como el ciudadano queda en una posición cuanto menos de desventaja frente al poder producido en la organización.

La omnipresencia del procesamiento de datos en la red trae aparejado una consecuencia directa para el ciudadano. Esta es la falta de transparencia de los algoritmos utilizados, es decir, una asimetría de información. Dado que son propiedad intelectual de las

organizaciones no suelen abrirse al público fácilmente (Mortier et. al., 2014, Victorelli et al., 2020). Pero además su diseño complejo impide que sean fácilmente interpretables e incluso no se sabe que datos son los que utilizan. Surge, entonces, una imposibilidad del ciudadano de poder tener control sobre sus datos al ser procesados en la organización (Mortier et. al., 2014).

Las organizaciones en general justifican la necesidad de recopilar casi cualquier dato como necesario para tomar una decisión en favor del ciudadano (Gandy Jr, 2021). Pero cierta parte de la población manifiesta que trae aparejado consecuencias sobre sus vidas. Existe una percepción de que la utilización de la tecnología resulta cuanto menos invasiva, generándoles preocupación. Su correlato inmediato es la alteración de la dinámica de las relaciones sociales en la medida que modifican las formas de su comportamiento. En el contexto de la pandemia COVID-19[6], es posible encontrar un claro ejemplo.

En el escenario de la pandemia, la facilidad y rapidez de recopilación de datos para realizar proyecciones confluyó en medidas de confinamiento adoptadas por las Autoridades Nacionales de Argentina en marzo 2020[7]. La consecuencia inmediata ha sido el cambio en las conductas de la sociedad. Desde la modalidad de trabajo remoto hasta restricciones en un principio totales para circular por la vía pública. A su vez, se presentó la imposibilidad del titular de los datos (el ciudadano) de poder determinar si las inferencias realizadas con sus datos eran correctas o de corregir ciertos datos o de decidir si darse de baja o continuar siendo parte del proceso (Mortier et. al., 2014). Es en este sentido que puede interpretarse como el individuo ve su capacidad de acción limitado.

Por ello, resulta importante no solo tener en cuenta a las personas que acceden y usan los datos. Si no que también debe incluirse a aquellos afectados por su uso (Victorelli et. al., 2020) a la hora de diseñar los sistemas de datos dentro de un contexto organizacional. Entre otras cosas porque la privacidad de los individuos es considera un valor social que puede verse corrompido en el contexto de la sociedad de la información. Particularmente, la privacidad asociada a datos es entendida como el derecho que poseen los individuos a saber que se hace con sus datos para tener cierto control. Si bien esta problemática será abordada

[6] El 12 de febrero de 2020, un virus de rápida expansión se lo denomina oficialmente enfermedad infecciosa bajo el nombre de Coronavirus 2019 (COVID-19). A partir de su rápida expansión, el 11 de marzo de 2020 la OMS lo declara una pandemia.
[7] Mediante el decreto 297/2020 emitido el 19 de marzo de 2020, las Autoridades Nacionales de Argentina, decretan el Aislamiento Social Preventivo y Obligatorio a fin de proteger la salud pública.

con mayor profundidad en el capítulo 2 del presente trabajo, es en este sentido que surge la necesidad de adecuar las normas sociales más allá de cualquier regulación o legislación establecida para dar respuesta a las consecuencias asociadas (Mortier et. al., 2014, Victorelli et. al., 2020).

Desde esta perspectiva y en el contexto de la sociedad de la información, el dato, puede ser visto como un elemento conductual. Se constituye en el elemento por medio del cual las conductas humanas pueden ser modificadas (Zuboff, 2015) como en el ejemplo anteriormente mencionado. De este modo, el accionar de las organizaciones podrá conformar un "poder de vigilancia" (Zuboff, 2015, 2019) en la medida que modifica comportamientos humanos donde la tecnología actúa como un "imperativo tecnológico" (Gandy Jr., 2021). Esto es, el poder de acción superior de la organización no reside en la tecnología en si misma sino en las personas que la utilizan. Los actores que poseen más información pueden utilizarla en la interacción con los ciudadanos restringiendo sus opciones o la manera en que entienden deben accionar (Zuboff, 2019, Gandy Jr., 2021).

Bajo este escenario posible, las organizaciones al obtener más información sobre el ciudadano generan una asimetría de conocimiento que le otorga mayor poder. A su vez, se configura un un escenario de nula incertidumbre para estas y sin mediar contrato alguno (Zuboff, 2015, 2019). Esto es posible en la medida que los individuos son inducidos a ceder sus datos para poder satisfacer sus necesidades en este nuevo contexto. En este sentido, es notable el volumen de información que manejan las organizaciones y el potencial predictivo que poseen a través del uso de tecnología.

De este modo, de la interacción entre los individuos y las organizaciones en un entorno digital tecnológico, es que los primeros dejan huellas digitales (rastros) como consecuencia de su accionar. Tales rastros son codificados a través de la tecnología y luego convertidos en datos. Al aplicar conocimiento mediante un proceso algorítmico – proceso que es desconocido por los individuos–, las organizaciones logran extraer información de estos. Si bien esta puede ser utilizada para favorecerlos, existe un riesgo de que pueda ser manipulada en su perjuicio. Considerando que entre el diferente tipo de organizaciones se encuentran las estatales, se pone en duda el rol del Estado como garante de los valores públicos.

Por esta razón, se requiere también tener en cuenta las consecuencias que genera la omnipresencia del procesamiento de datos sobre los individuos. Para poder lograrlo, resulta

necesario colocar al ciudadano en el centro de la circulación de datos y considerar el impacto causado. Así será posible observar como el uso de datos en la red puede modificar los comportamientos humanos generándose una disparidad de poder sustentada en una asimetría de conocimiento. Esto permitirá allanar el camino hacia sentar las bases para un uso responsable de la información.

1.1.3 El poder producido en base a información

El poder que construyen las organizaciones en base a su capacidad superior de generar información puede encuadrarse en la propuesta de Randall Bartllet. El autor propone que el poder es "la capacidad de un actor para alterar las decisiones tomadas por otro actor en relación con las elecciones que se habrían hecho si el primer actor no hubiera existido o actuado" (Bartlett, 2006, pág. 42). En la red, son las organizaciones las que poseen los recursos tecnológicos y especializados para la obtención de datos y generación de información. Al obtener un mayor conocimiento, surge la posibilidad de que puedan alterar la decisión de un individuo. A partir de ello es que se las puede posicionar en la red como el actor que posee mayor poder de acción.

La obtención de los datos surge de una relación de intercambio llevado a cabo en la red. Por un lado, los individuos reciben una contraprestación que puede ser un bien o servicio para satisfacer su necesidad. Por el otro, las organizaciones reciben un bien que son sus datos a partir del cual obtienen un rédito. Hasta aquí podría pensarse que se trata de un típico intercambio en un mercado. Pero también surge la posibilidad de que estas últimas manipulen el recurso obtenido para modificar la decisión del individuo. Por ello resulta importante tener en cuenta las principales particularidades de cómo es producida la información en la organización.

El recurso primario en la producción de información son los datos. A diferencia de los bienes primarios de producción –en términos clásicos, los obtenidos de la naturaleza principalmente–, el dato posee la particularidad de ser un bien intangible (Haskel y Westlake, 2017). A su vez el capital esencial utilizado para llevar a cabo la generación de información en base a este también es intangible. Este último, abarca tanto el conocimiento especializado como tecnologías digitales de *Big Data*. El rol de este tipo de tecnología juega un papel central, ya que sin estas no sería posible ni la obtención ni el procesamiento de datos.

Las tecnologías de *Big Data* no solo abarcan bienes de capital físico (computadoras, servidores, entre otros), si no también abarcan la inversión en capital digital, principalmente de *software*, procesos, conocimiento, ideas. El conjunto de todos ellos hace a la conformación de la tecnología digital o tecnología de la información (Shapiro y Varían, 1998, Haskel y Westlake, 2017) o tecnología intelectual como fue definida en el primer apartado. Pero, además, la organización tiene la capacidad de generar una "sinergia" (Haskel y Westlake, 2017) entre todos esos elementos. Al captar a los especialistas, logra conseguir que se aplique conocimiento a datos a través de tecnología para que finalmente se produzca información.

Por un lado, tales características implican que la mayor capacidad de producir información recae en manos de la organización. Si bien existe un costo inicial para producirla, el costo de reproducirla es muy bajo por lo que existen costos marginales tendientes a cero (Shapiro y Varían, 1998). Con ello se modifica la forma en que se produce la valorización ya que el precio no vendrá determinado por el costo – en términos de la teoría económica clásica de formación de precios[8]–. A su vez, el gran flujo de información que circula rápidamente en la red a través de internet le otorga la posibilidad de construir una mejor personalización para generar estrategias de acción más eficientes y diferenciales (Shapiro y Varían, 1998; Varían, 2010).

Por otro lado, implica un cambio en la manera en que se da la interacción con los individuos. Es decir, ya no se requiere de un cara a cara para llevarse a cabo un intercambio, sino que se sustenta en una relación de confianza mediada por computadoras (Varían, 2010; Haskel y Westlake, 2017). Esta es la que depositan los individuos en las organizaciones. Con ello, la reputación organizacional cobra mayor relevancia (Stiglitz, 2000). En la medida que se produce información inexacta conducirá a una toma de decisiones errónea (Bartlett, 2006) que puede derivar en la pérdida de tal confianza.

Desde esta perspectiva puede considerarse que el entorno digital donde se produce el intercambio se constituye en un mercado. A diferencia del mercado en términos de la economía clásica, este posee la particularidad de conformarse por un ensamblaje entre los actores y la tecnología en la red. Si bien esta noción será desarrollada con mayor especificación en el apartado 1.3., resulta importante considerar la omnipresencia del

[8] Para la cual a partir de la interacción entre la demanda agregada y la oferta surge la definición de un precio (Coumot, 1838).

proceso de generación de información. Las organizaciones logran construir un poder de acción superior al del individuo sustentado en información. Esta es extraída de los datos que captura en este contexto tecnológico en particular. A su vez, sólo lo puede realizar a través de aplicar conocimiento a datos mediante el uso de tecnologías digitales.

El proceso de generación de información ocurre de manera no visible para el individuo. Frente a la pérdida de control que estos últimos sufren, es posible la construcción del poder superior por parte de las organizaciones. Autores clásicos como Bartllet, sostienen que "siempre que el conocimiento es un bien escaso, confiere poder a sus poseedores" (Bartlett, 2006, pág. 101). A partir de este tipo de justificativo, el poder es concebido como una externalidad negativa más en el contexto de una relación de intercambio.

En cambio, autores como Zuboff (2015), Ghandy Jr. (2021), sostienen que el poder concentrado en las organizaciones deriva en una limitación concreta de la capacidad de acción del individuo. La organización tiene el poder de utilizar la información generada como medio (Zuboff, 2015) para modificar el escenario de acciones. A su vez, esto le permite crear una forma de vigilancia. También, puede decidir qué información muestra o no (Gandy Jr, 2021). Frente a estas opciones, la posibilidad de decidir del individuo es cuanto menos escasa. Pero también, podría generarse cierta manipulación para su conveniencia.

De aquí que puede resultar considerable la concepción del poder producido (Latour, 2005). En la red digital, el poder ya no surge de la acumulación de capital como sostiene la teoría económica clásica. Sino que es producido y sostenido en base a información. Pero, además, el sistema tecnológico digital se convierte en un recurso central para que esto suceda. Ya no solo debe ser entendido como un recurso productivo más sino como un factor esencial que contribuye a la conformación de esta nueva forma de poder.

A partir de lo expuesto en el párrafo anterior, el tipo de organización que se configura es aquella que basa sus decisiones en información. No solo se trata de las creadas desde el inicio como una de este tipo. También involucra a aquellas que tenían una estructura tradicional y se han reconvertido en parte o en su totalidad a una de estas. Entre estas últimas se encuentran las organizaciones públicas del estado. La adopción de las nuevas tecnologías del *Big Data* para el procesamiento de datos juega un rol central en este sentido y permitirá enmarcarlas dentro de lo que Bruno Latour (1987; 2005) define como centro de cálculo.

1.2 Organizaciones públicas que basan sus decisiones en datos

En la era de la sociedad de información, las organizaciones típicas poseen la particular características de adoptar decisiones en base a evidencia (McAfee y Brynjolfsson, 2012). Son aquellas que sustentan sus decisiones en información la cual surge del procesamiento datos mediante tecnología y aplicación de conocimiento. Las estructuras productivas tradicionales ya no resultarán adecuadas. Se requiere de un proceso de adaptación para encontrar patrones en los datos y traducirlos en información útil. Además, dada la particularidad de intangibilidad de los recursos necesarios (datos, tecnología, conocimiento), se comienza a transitar una revolución organizacional impulsada por los datos. (Milgrom, 1992; McAfee y Brynjolfsson, 2012; Schmarzo, 2013; Haskel y Westlake, 2017).

Como fue definido en el apartado1.1.2, los datos son producidos por los individuos en su accionar en la sociedad en red. En este sentido, se trata de una unidad mínima de información que se encuentra dispersa. Ante ello se requerirá de una coordinación adecuada para recolectarlos, procesarlos y generar valor agregado (McAfee y Brynjolfsson, 2012) en la organización. A su vez, esto expone la complejidad a la cual deben enfrentarse las organizaciones a la hora de trabajar con datos.

El impacto causado por esta nueva forma de producir valor agregado ha alcanzado tanto a organizaciones de carácter privado como público. En el primer caso, se trata principalmente de aquellas conocidas como organizaciones de plataformas. Su principal característica definitoria es que son una red online de interacciones (Zysman y Kenney, 2018) – entre las que se encuentran Mercado Libre, Amazon, Google, entre muchas otras–. Actúan de intermediarias entre los actores de la sociedad poniendo a su servicio una infraestructura tecnológica digital para que se produzca un intercambio. Específicamente, se trata de negocios basados en información en red (Moazed y Johnson, 2016) donde la interacción es mediada por una computadora (Varian, 2010).

En cambio, las organizaciones estatales no tienen un fin comercial. Su principal objetivo es realizar políticas públicas para resolver problemas grandes o graves que el propio mercado no puede resolver. Como resultado la sociedad recibe un servicio o un bien público (Bryson et al., 2015; Haskel y Westlake, 2017). Es decir, son generadoras de valor público el cual se define como todo asunto de interés público (Naser, 2021). Abarca desde bienes y servicios que satisfagan una necesidad de los diferentes actores de la sociedad hasta el uso legítimo de recursos para lograr diversos propósitos públicos. Si bien históricamente las

organizaciones públicas del estado han utilizado datos para el ejercicio de sus funciones, también han sido impactadas por la transformación digital. Esto ha dado lugar al surgimiento de lo que se conoce como gobierno digital.

El gobierno digital se define como el ecosistema conformado por los diversos actores sociales (organizaciones de gobierno, ciudadanos, organizaciones privadas) y el uso de tecnologías digitales como parte integral de la estrategia para la creación de valor público (OCDE, 2014). A partir de ello se considera al Estado como una organización en si misma que obtiene datos al interactuar en un entorno digital con los ciudadanos. Su gestión adecuada a través del uso de tecnología le permitirá obtener información que será utilizada como evidencia para la toma de decisiones. Esto es, le permite diseñar políticas públicas más eficientes y eficaces en pos de la generación de valor público.

Al mismo tiempo, resulta importante realizar una distinción según el fin para el cual los datos son utilizados. Por un lado, existe una recopilación de datos a través de por ejemplo un censo nacional. Esto permite la elaboración de estadísticas agregadas para obtener cierto conocimiento sobre una población (Gandy Jr, 2021). Por el otro, cada organización pública que conforma el sistema administrativo del estado nacional recopila datos de los ciudadanos con el propósito de llevar adelante una tarea burocrática particular (Gandy Jr, 2021). En Argentina, por ejemplo, dentro del primer caso se encuentra el Instituto Nacional de Estadísticas y Censos (INDEC); mientras que en el segundo caso es posible identificar a la Administración Federal de Ingresos Públicos (AFIP), entre otras.

Por otra parte, también se produce una trasferencia de datos entre las distintas organizaciones estatales (Gandy Jr, 2021). En este proceso no hay participación ciudadana donde el individuo pueda ejercer cierto control. Es un proceso que se da de manera invisible ya que desconocen cómo, cuándo y porqué se lleva a cabo. Como fue mencionado en el apartado anterior, queda expuesto el poder diferencial que adquiere el propio Estado y la situación de vulnerabilidad en la que puede quedar el ciudadano. También, puede surgir el riesgo de que las representaciones creadas en base a los datos sean erróneas o sesgadas. En consecuencia, los beneficios o servicios que los individuos recibirán (Gandy Jr, 2021) pueden resultar inapropiados o insuficientes.

A partir del contexto organizacional que se configura, en este apartado se centra el análisis en el rol que desempeña la tecnología. Para ello, en una primera sección, se desarrolla el

concepto de tecnologías de *Big Data.* Su abordaje conceptual permitirá especificar acerca de la capacidad de cálculo derivada lo que le da forma a las organizaciones como un centro de cálculo. Luego se especifica acerca de del papel de la gestión de datos como esencial para alcanzar la generación de valor. Finalmente, se establece como se desarrolla la toma de decisiones en base a estos. De este modo, podrá caracterizarse a una organización que basa sus decisiones en datos.

1.2.1 Tecnologías de *Big Data* como factor clave de la capacidad de cálculo en organizaciones

Desde comienzo del siglo XXI, el impacto causado por los nuevos desarrollos tecnológicos ha dado lugar al surgimiento de un conjunto de tecnologías digitales denominadas *Big Data*. Ante la disponibilidad de grandes volúmenes de datos, el término se basa en la idea de que no es posible manejar aquello que no se puede medir (McAfee y Brynjolfsson, 2012; Schmarzo, 2013). Pero también se refiere a la necesidad de diseñar un proceso adecuado para obtener mejores resultados que derivarán en un mayor conocimiento para sustentar decisiones.

El rol de las tecnologías de *Big Data* en la organización no solo implica la implementación de métodos y técnicas mediante la utilización de lenguajes de programación o *Software*. Si no que también involucra la definición de una estrategia para llevar a cabo el procesamiento adecuado para lograr extraer información a partir de los datos (Constantiou y Kallinikos, 2015). En este sentido, el concepto evoluciona hacia un sistema de mayor complejidad.

El *Big Data* abarca un conjunto de tecnologías que fueron transformando de manera disruptiva a la sociedad en su conjunto. Se trata de un "ecosistema" (Schmarzo, 2013; Constantiou y Kallinikos, 2015; Kolanovic y Krishnamachari, 2017) que ha permitido manejar la complejidad de descubrir patrones en grandes y diversos volúmenes de datos para la generación de información valiosa. Si bien no existe una única definición del término, hay un consenso generalizado en la literatura sobre tres características principales: volumen, la velocidad y la variedad.

Por volumen se hace referencia a la tecnología necesaria para recolectar y almacenar grandes volúmenes de datos con el fin de procesarlos para transformarlos en información de utilidad. Esto encuentra su origen en la masividad del uso de dispositivos tecnológicos con conectividad por parte de los individuos. El crecimiento exponencial de los datos agrega

complejidad para su almacenamiento y procesamiento. Pero al mismo tiempo, el desarrollo tecnológico ha facilitado las herramientas para que las organizaciones encuentren una forma clara de explotarlos (McAfee y Brynjolfsson, 2012; Constantiou y Kallinikos, 2015; Eberendu, 2016; Kolanovic y Krishnamachari, 2017).

En cuanto a la velocidad, resulta ser en muchos casos más importante que el volumen (McAfee y Brynjolfsson, 2012) ya que se requerirá de que una organización cuente con la agilidad suficiente para captar los datos en tiempo real. En este sentido, los datos ya dejan ser un stock para ser un flujo constante (Eberendu, 2016). Implica un procesamiento diario y hasta en ocasiones por hora. De aquí que, el procesamiento requiere de cierta complejidad superior para lograr captar la mayor cantidad de tipos de datos provenientes de diferentes fuentes. A su vez, requiere de hacerlo con la mayor velocidad posible. De este modo podrán convertirlos en valor agregado para la toma de decisiones más inmediatas (Constantiou y Kallinikos, 2015; Kolanovic y Krishnamachari, 2017).

La variedad o diversidad de datos resultan de la existencia de plataformas digitales donde los individuos interactúan a diario. En este sentido se habla de cambios en las conductas de los individuos configurándose una nueva cultura social (Castells, 2002). Con ello ha surgido el diseño de procesos que estructuren, sinteticen y simplifiquen la información que pueda obtenerse. Pero al mismo tiempo, se presenta la posibilidad de contar con información sobre cualquier tema o actividad de interés del individuo (McAfee y Brynjolfsson, 2012; Constantiou y Kallinikos, 2015; Eberendu, 2016; Kolanovic y Krishnamachari, 2017). De este modo, el *Big Data*, refiere al proceso por medio del cual las organizaciones se apropian de los datos generados por los individuos al interactuar en red para convertirlos en información.

Autores como Gandy Jr. (2021) consideran que el ecosistema del *Big Data* otorga un poder superior a las organizaciones. Sostiene que les permite obtener mayor conocimiento sobre los individuos más allá de lo que ellos mismos puedan saber o están dispuesto a mostrar. Las organizaciones y en particular las organizaciones públicas del estado adquieren un nivel mayor de conocimiento sobre los ciudadanos. Si bien puede ser utilizado con un fin bien intencionado, surge el riesgo de que suceda lo contrario.

De aquí que alrededor de la disrupción del *Big Data* en la sociedad también se han generado diferentes posturas críticas. La recopilación masiva de datos en la red a través de tecnologías

digitales en una primera etapa ha sido valorada por su potencial para la generación de valor. Pero, ya en la década del 2010, se comienza a plantear cuestionamientos alrededor de su uso. Un caso emblema que puso en jaque la utilización ilimitada de esta tecnología, fue lo sucedido con *Cambridge Analytica* para la campaña electoral de Donald Trump en 2016. Si bien este caso será abordado en el capítulo 2, resulta en un claro ejemplo del poder de este tipo de tecnología para procesar datos sin el consentimiento de sus titulares. A su vez, de cómo la información generada puede modificar decisiones humanas[9].

El riesgo asociado la utilización de tecnología para el procesamiento de datos comienza a ser más evidente. En ocasiones la información obtenida puede ser manipulada para perjudicar a los individuos o para inducirlos a tomar una decisión no beneficiosa para sí mismo. En otros casos, la captura de datos sin el consentimiento de su titular puede derivar en la violación del derecho a su privacidad (AEPD, 2020). Particularmente, este último caso es de mayor interés en este trabajo y será abordado en el capítulo 2.

Ahora bien, en el contexto estatal, las tecnologías del *Big Data*, sin duda, han permitido ofrecer servicios con eficacia, eficiencia, disponibilidad, interoperabilidad y con racionalización de los recursos, entre otros beneficios (AEPD, 2020). Las organizaciones públicas del estado han encontrado en estas un medio para la aplicación de conocimiento para la ejecución de cálculos sobre los datos. El resultado obtenido, les ha facilitado el diseño de políticas públicas de una manera más rápida y ágil. El valor agregado conseguido constituye la evidencia sobre la cual sustentan la toma de decisiones.

Al decir que las organizaciones realizan cálculos, implica que existe una necesidad preexistente de hacer inteligible los datos (Rivoir y Morales, 2019). Existe una necesidad de codificar los datos para luego generar información que pueda ser comprendida. Esto ha sido posible como resultado de aplicar un proceso de algorítmico a través de las tecnologías digitales. El fin último es poder interpretar las acciones (rastros) de los individuos en un entorno digital y convertir a la información en valor agregado para la toma de decisiones.

Las tecnologías de *Big Data* son las que permiten llevar adelante el proceso de conversión de datos en información (Constantiou y Kallinikos, 2015; Kolanovic y Krishnamachari, 2017). De este modo, la realización de cálculos puede ser concebida como un proceso algorítmico a través del cual los datos son convertidos en valor. A partir de ello, es posible

[9] Para el ejemplo en particular, como puede manipularse el voto de un electorado.

convertir lo intangible en informes de resultados, gráficos, tablas de datos, entre otros (Haskel y Westlake, 2017).

La configuración de una organización que basa sus decisiones en datos, a diferencia de lo que sostiene Varían (2010) no solo se trata de una simple interacción mediada por una computadora. Si no que es el proceso por medio del cual se obtienen datos, se genera más información y se define una acción a seguir. En este sentido es que la organización puede ser interpretada como un centro de cálculo (Latour 1987; Callon y Muniesa, 2003; García Fronti y Herrera, 2021). Aquel espacio donde el conocimiento especializado y la tecnología facilitadora de cálculos interactúan para desarrollar una capacidad de acción que no es conocida por el ciudadano común. Siguiendo a los autores Callon y Muniesa (2003), este proceso de cálculo puede concebirse de modo general en tres pasos.

En primer lugar, las entidades deben ser aisladas, es decir, los datos deben ser recolectados y almacenados. A partir de ello comenzarán a circular en el espacio del centro para ser transformados a través de un cálculo. Para ello, en segundo lugar, será necesario asociarlos con otras entidades. Estas vendrán dadas por un conocimiento previo a través de especialistas y en conjunto serán sometidas a un proceso de conversión a través de tecnología. Finalmente, se podrá extraer una nueva entidad que es un resultado, siendo el valor agregado generado y que será utilizado para tomar una decisión.

El proceso de cálculo permite comprender que la capacidad de acción de la organización se configura a partir de un "híbrido colectivo" (Callon y Muniesa, 2003; García Fronti y Herrera, 2021). La acción conjunta de la tecnología (no humana) y los individuos (humanos) que poseen conocimiento especializado, es lo que les permite obtener información a partir de los datos. Por ejemplo, para que un analista pueda llevar a cabo un informe analítico en base a datos, estos deben previamente ser obtenidos. Posteriormente deberán ser procesados a través de la tecnología y luego transformados mediante un proceso algorítmico para lograr construir información. El conocimiento especializado, también se lo requiere para elaborar un análisis de la información obtenida de modo tal que sea posible la construcción de valor agregado. Por lo tanto, no solo se trata de cálculos mentales realizados por individuos especializados, sino que son integrados con el uso de la tecnología para generar información o valor en base a datos.

De esta manera, en el contexto de la sociedad de información, las organizaciones son un actor en la red. Entre estas también se encuentran las públicas del Estado. Al poseer una capacidad de cálculo superior pueden interpretarse como un centro de cálculo. Esta capacidad la adquiere a partir de apropiarse de los datos generados por los ciudadanos, transformándolos mediante los recursos tecnológicos y humanos que dispone en información. La información generada les permite obtener mayor conocimiento. Ahora bien, para completar este proceso con éxito se requerirá de un sistema de gestión de datos integrado. Este les permitirá obtener información de calidad, verídica y relevante.

1.2.2 La gestión de datos frente al impacto del *Big Data*

En el proceso de cálculo descripto en el apartado anterior, los datos recolectados deberán asociarse con otras definiciones transversales de la organización. En general, estas son propias al objetivo particular que posee. Además, facilitan la convivencia de diferentes sistemas y procesos para la construcción de información. Por esto, la claridad de su definición es un objetivo central de un plan de gestión de datos, ya que permitirán integrar, analizar y extraer el valor que contienen los datos (Cleven y Wortmann, 2010). De esta manera la gestión de datos toma un rol central, dado que se trata de un proceso a partir del cual la organización gestionará la información como un recurso integrado (McKeen y Smith, 2007).

Para garantizar la confiabilidad en los datos surge la necesidad de establecer reglas y políticas internas para el manejo de datos en la organización. En este sentido, las entidades centrales (o datos maestros) se convierten en un activo clave. Son aquellas definiciones globales pero propias que responden al fin último de funcionamiento de los procesos, permitiendo asociar cada dato a una entidad (McKeen y Smith, 2007; Cleven y Wortmann, 2010). Por consiguiente, la falta de una adecuada gestión de estos puede derivar en el surgimiento de diferentes problemas. Por ejemplo, fallas en el funcionamiento de procesos hasta una mala toma de decisiones por generación de información incorrecta.

Desde una perspectiva técnica, el rol principal de la gestión de datos es la gestión del riesgo técnico asociado a la estructura, la arquitectura, el gobierno, el proceso y la calidad de definiciones centrales (Cleven y Wortmann, 2010). La estructura refiere a la necesidad establecer un acuerdo sobre cada definición que satisfaga las necesidades de cada área de la organización. A partir de ello se establecerá una estructura relacional. La arquitectura implica diseñar sistemas adecuados que den soporte al ciclo de vida de los datos. En cuanto

a la gobernanza de datos, implica la definición de roles y responsabilidades sobre el uso de los datos dentro de las diversas áreas de la organización. Para ello será necesario definir un proceso adecuado sobre la manera en que los datos deben ser creados, usados, mantenidos y almacenados. Finalmente, la calidad será alcanzada como consecuencia de la implementación correcta de las cuatro fases anteriores.

Para garantizar un resultado exitoso en la calidad de datos a su vez se requiere de un proceso interactivo continuo (Cleven y Wortmann, 2010). Este consta de tres partes principales. En una etapa inicial será necesario realizar un análisis para identificar cuáles serán las entidades centrales de la organización. En una segunda etapa se definirá los conceptos asociados que serán transversales a toda la organización. Y, finalmente, se requerirá de una etapa de control y revisión continua que permita garantizar la sincronía de los datos o identificar necesidades nuevas. De esta manera, se logrará un sistema de datos consolidado.

Esto último expone que para generar información no resulta suficiente solo considerar el aspecto de gestión técnica de la tecnología. Sino que el aspecto humano para la interpretación y análisis también resulta clave en la generación de conocimiento. En este sentido, se requerirá de individuos con liderazgo que sean capaces de hacer las preguntas correctas para encontrar el éxito de los resultados obtenidos. No se trata solo del uso de la tecnología del *Big Data* en sí, sino de las personas que la utilizan (McAfee y Brynjolfsson, 2012; Gandy Jr, 2021). De aquí que también surgirá la creación de una cultura en base a datos en la organización (McAfee y Brynjolfsson, 2012).

Ahora bien, entre los datos recolectados se encuentran aquellos que son personales de cada individuo (Gandy Jr, 2021). Si el destino de uso es mal intencionado, podría afectar la integridad del individuo titular. Ante ello, es posible notar que la responsabilidad de los actores intervinientes jugará un rol fundamental. No solo se trata de asegurar el aspecto técnico de gestión sino también el sentido de compromiso frente al recurso utilizado. A partir de aquí puede extenderse el concepto técnico de diseño del sistema hacia uno que abarca también las responsabilidades involucradas.

Al colocar al ciudadano en el centro del diseño de un sistema de gestión de datos, se podrá evaluar los riesgos asociados. Tal construcción permitirá adoptar una actitud proactiva antes que reactiva. Se podrá actuar en la prevención de los futuros posibles daños antes que accionar una vez que ya sucedieron. De esta manera, las responsabilidades serán asignadas

por defecto a quienes se encuentren involucrados (AEPD, 2019, 2020). De esta manera, el diseño de una estrategia de gestión podrá incorporar también la protección de datos personales. Sobre este punto se profundizará en el capítulo 2.

Así, la organización logrará establecer un sistema de datos consolidado que le permitirá reducir los riesgos asociados y generar información adecuada. Al llevar adelante una gestión de datos eficiente mediante la aplicación de tecnológicas de *Big Data*, podrá diseñar estrategias de acción óptimas. En el caso de las organizaciones públicas le permitirá elaborar políticas públicas eficientes y eficaces, aportando valor en la sociedad. En consecuencia, la extracción de conocimiento podrá ser utilizado para llevar adelante una toma de decisiones basada en evidencia en un marco responsable.

1.2.3 La producción de conocimiento para la toma de decisiones

A diferencia del pasado, cuando los científicos generaban conocimiento en abstracto, ahora cuentan con mayores facilidades para hacer ciencia empírica (Rhodes y Lancaster, 2019). La disponibilidad de grandes volúmenes de datos y de tecnologías del *Big Data* en los centros de cálculo, les facilita una red de interacción para la elaboración de conocimiento. Podría decirse que, en la actualidad, la ciencia deja de ser un conjunto de meras posibilidades teóricas para materializarse en base a evidencia, en información.

En el contexto de la sociedad de la información, la ciencia es vista como la ciencia que participa en la materialidad relacional de implementar y evidenciar (Rhodes y Lancaster, 2019). Es decir, las proyecciones producen conocimiento, ofreciendo una sensación de seguridad a través del cálculo (Rhodes, et al., 2020). En este sentido, podría decirse que el conocimiento científico deja de ser meras posibilidades teóricas para concretarse en toma de decisiones que afectan las acciones en el presente. A partir de ello, se configura una nueva disciplina científica que es la ciencia de datos. Su fin último es hacer más eficiente la toma de decisiones en la organización mediante la comprensión de los fenómenos a través de un análisis automatizado de datos (Provost & Fawcett, 2013). Se trata de un proceso que involucra no solo la etapa de análisis, sino que se fundamenta en un proceder fundamentado en principalmente dos conceptos centrales.

El primer concepto fundamental es que se trata de la extracción o descubrimiento de conocimiento útil para resolver un problema. En relación con esto existe una responsabilidad que recae en la figura del cientista de datos. Se trata de contar con una capacidad de

comprender las necesidades y el contexto alrededor del problema objetivo. Pero también, de poder hacer una evaluación de los recursos disponibles (Provost & Fawcett, 2013, Schmarzo 2013) para materializar una posible solución. Las tecnologías del *Big Data* jugarán un rol importante en este sentido y se convierten en el segundo fundamento central en el contexto de la ciencia de datos. Estas serán las que permitan aplicar una transformación de los datos para obtener datos informativos (Provost & Fawcett, 2013, Schmarzo 2013). Tal potencial evidencia el poder de instrumentación que brinda el uso de tecnología (Zuboff, 2019) así como la necesidad de gestionar en un marco de responsabilidad como se mencionó en el apartado anterior.

Como resultado de la instrumentación de dichas tecnologías, las decisiones ya no son tomadas en base a intuiciones. Sino que se fundamentan en información obtenida a partir de aplicar conocimiento científico a datos mediante tecnología. El resultado obtenido podrá generalizarse para establecer relaciones de causa-efecto (Rousseau, 2006). En el caso del Estado, podrá adoptar medidas para la ejecución de políticas públicas. La medida de aislamiento social impuesta en Argentina en pandemia mencionada en un apartado anterior constituye un claro ejemplo. Particularmente, la creación de un aplicativo móvil denominado CuidAR[10] le permitió al Estado obtener más datos para realizar políticas públicas de control de la pandemia (Salaberry y Herrera, 2021). De esta manera, el poder de acción de las proyecciones obtenidas no sólo se traduce en ciencia y política sino en acciones ciudadanas y formas de vida social (Rhodes, et al., 2020). A su vez, permite notar como la capacidad de cálculo es utilizada para el ejercicio del poder de control anticipatorio basado en evidencia a través de tecnología (Zuboff, 2019; Rhodes, et al., 2020).

La ejecución práctica del conocimiento de manera anticipada permite a las organizaciones públicas convertir el futuro incierto en un objeto de estudio en el presente. Esto facilita su intervención a través de políticas públicas. La producción de información a través de tecnología sobre lo que aún no ha sucedido les permite reducir la complejidad social a un problema concreto para ejercer una "gobernanza anticipatoria" (Guston, 2014; Aykut, Demortain, & Benbouzid, 2019). En un caso de pandemia puede considerarse necesario para garantizar la salud de una población en general. Pero no debería perderse de vista que la

[10] Acerca de la App CuidAR: https://www.argentina.gob.ar/jefatura/innovacion-publica/acciones-coronavirus/aplicacion-y-tableros-de-gestion

recopilación masiva de datos se trata de un proceso omnipresente por medio del cual se configura un poder de acción superior en la organización.

Del mismo modo que en cualquier investigación científica para la generación de conocimiento se exige una cuota de moralidad o responsabilidad en el uso de los datos, surge la posibilidad de preguntarse si en este proceder organizacional existe algún riesgo asociado. Particularmente porque entre los datos recopilados y procesados se encuentran los personales de los ciudadanos. Sumado al poder predictivo de las tecnologías de *Big Data*, se encuentran cuatro acciones susceptibles de considerarse riesgosas.

La primera acción riesgosa es la reutilización de los datos. Particularmente implica la responsabilidad sobre poder disponer de los datos y que sean utilizados únicamente para el fin por el cual fueron recolectados. Esto último, da lugar a la segunda acción riesgosa que es la posibilidad de readaptarlos para otros fines por el potencial que brindan las tecnologías. Una tercera acción riesgosa es la posibilidad de recombinar los datos recolectados con otros generando un conocimiento aún más específico o amplio de los individuos. Por último, también surge la posibilidad de reanalizar los datos por fuera del alcance de la imaginación de su titular. Estas cuatro actividades riesgosas derivadas del uso de datos a través de tecnologías de *Big Data* son lo que Steinmann, Matei y Collmann (2016) definen como 4R.

En consecuencia, la responsabilidad en el uso de datos para la generación de conocimiento no solo se refiere a la capacidad de llevar adelante el proceso, sino también la de considerar principios para un accionar responsable. Por esta razón, además de gestionar eficientemente los datos, las organizaciones requieren contar con un marco para llevar adelante una gobernanza responsable de aquellos. En el caso particular de las organizaciones públicas, la existencia de dicho marco le permitirá entre otras cosas generar una mayor confianza con la ciudadanía. A su vez les facilitará aportar más valor a la sociedad a través del diseño y ejecución de políticas públicas eficientes y eficaces.

De este modo, las organizaciones públicas que basan sus decisiones en datos en el contexto de la sociedad de la información pueden ser concebidas como un centro de cálculo. Esto implica que son generadoras de conocimiento. Para lograrlo interrelacionan saber científico con tecnologías de *Big Data* aplicándolos a los datos. Estos les pertenecen a los individuos y son capturados de su interacción en la red. Frente a los riesgos que pueden suscitarse podrán diseñar un sistema de gestión eficiente de los recursos en un marco de gobernanza responsable.

Pero al mismo tiempo, el proceso de obtención de datos en red les otorga un poder superior a las organizaciones públicas. Por un lado, les facilita el ejercicio del control de escenarios futuros inciertos a través de intervenir con políticas públicas (gobernanza anticipatoria). Por el otro, el poder de cálculo superior con el cuentan a través del uso de tecnologías de *Big Data* les permite modificar conductas humanas.

Dada esta configuración, lo primero que sucede es la interacción con los ciudadanos para hacer posible la obtención de datos. Las organizaciones son las propias creadoras y controladoras del espacio donde ello sucede. Se trata de entornos digitales que reciben el nombre de plataforma. En este espacio, logran capturar los datos de los ciudadanos que luego mediante la aplicación de conocimiento a través de las tecnologías de *Big Data* transforman en información. El cuestionamiento posible de plantear es acerca de cómo esta es utilizada. Si bien puede ser utilizada como evidencia para la toma de decisiones, en ocasiones puede ser empleada como medio para modificar las acciones posibles de los individuos. No solo dentro el mismo espacio digital, sino también en su accionar diario en la sociedad. De este modo, es posible interpretar que las organizaciones logran producir un poder mayor de acción el cual se sostiene en base a información y no en la acumulación de capital.

A su vez, la plataforma al ser un espacio digital de intercambio puede ser interpretado como un mercado, particularmente de datos. Desde esta perspectiva, es posible establecer que se produce una asimetría de conocimiento. Pero surge la posibilidad de cuestionar acerca de cómo influye en la construcción del poder de la organización. Para poder ahondar en este cuestionamiento en el siguiente apartado se profundiza desde la concepción del mercado (de datos). Al comprender como se produce el valor en él, luego se aborda el impacto de la asimetría de conocimiento generada como parte constitutiva del poder de acción.

1.3 El poder de las organizaciones en el mercado de datos

La plataforma desarrollada por las organizaciones es el espacio digital donde se produce una interacción con los ciudadanos para realizar un intercambio. Mientras las primeras capturan los datos pertenecientes a los segundos, los ciudadanos reciben un bien o servicio. En este sentido es que la estructura de esta red digital puede ser concebida como un mercado particularmente de datos. En el es donde se produce una asimetría de conocimiento que es el origen causal de la concentración de poder en manos de las organizaciones y limita el accionar ciudadano. En este contexto, el objetivo de esta sección es abordar la conformación

de tal mercado y sus particularidades, así como también especificar acerca de la asimetría producida en tanto sustento del poder de acción superior producido en la organización.

El mercado de datos puede ser concebido como una red, específicamente una red sociotécnica (Caliskan y Callon, 2010, García Fronti y Herrera, 2021). En esta los agentes que la conforman ya no son meros agentes de intercambio como se propone en la perspectiva económica clásica, sino que son agentes sociotécnicos dotados de capacidad de cálculo. El principal agente son las organizaciones. Estas producen valor en base a datos a través del uso intensivo de tecnología. Los datos son obtenidos a partir de interactuar con los individuos en un espacio digital de intercambio (plataforma) que es creado y controlado por las propias organizaciones. Las tecnologías de *Big Data* son las que les facilitan la realización de los cálculos y amplían su poder de acción. Como contrapartida surge la conformación de una asimetría de conocimiento. Esta ocasiona una disputa de poder en la cual las organizaciones logran concentrar la mayor parte.

Además, sucede que la información elaborada en base a cálculos sobre los datos las organizaciones los utilizan como medio para modificar decisiones ciudadanas. Esto también permite interpretar a la red de interacción digital como un mercado de datos conductuales (Zuboff, 2019). La manera de llevarlo a cabo es a través de la modificación del escenario de opciones posibles entre las cuales podrán elegir los ciudadanos. El instrumento mediante el cual lo realizan son las tecnologías *Big Data.* Estas son las que brindan la posibilidad de convertir al dato en un predictivo de conductas humanas futuras lo que las constituye en un instrumento de poder. De aquí que las organizaciones poseen el control en este espacio digital ya que son ellas quienes crean estas plataformas.

En el caso particular de las organizaciones públicas del Estado, encuentran un incentivo puntual para desarrollar la red digital que es la recopilación de datos (Mazzucato, Entsminger y Kattel, 2020) de los ciudadanos. Estos poseen la cualidad de ser generadores de valor y son utilizados para intervenir mediante la realización de políticas públicas. Para ello ponen a disposición de la sociedad una red digital denominada gobierno electrónico. Los individuos, que poseen necesidades particulares están dispuestos a ceder sus datos para poder satisfacerlas (Zuboff, 2015; 2019; Ghandy Jr, 2021). A cambio reciben una contraprestación que puede ser un servicio público o un bien público.

Frente a la conformación de este mercado particular, resulta necesario profundizar acerca de su conformación y funcionamiento para finalmente comprender la asimetría de

conocimiento generada. Ello en la medida que sustenta la producción de poder en manos de las organizaciones. En la primera sección, se desarrolla una explicación de la constitución del mercado de datos. Luego, se pone de relieve como los datos personales son usados como medios para ejercer el poder de control del Estado y también para crear valor. Finalmente, se aborda la asimetría de información producida que sustenta la desigualdad de poder acción entre los agentes que intervienen.

1.3.1 El mercado como una red sociotécnica

Desde la perspectiva económica clásica, el mercado se define como un espacio de intercambio entre oferentes y demandantes (Shapiro y Varían, 1998; Stiglitz 2000, Bartlett, 2006). Pero con la disrupción de la tecnología este evoluciona hacia un mercado digital. A diferencia del mercado tradicional ahora ya no se requiere de un espacio físico para poder transaccionar, sino de una plataforma digital (Täuscher & Laudien, 2018). Estas son creadas y controladas por las organizaciones y su característica definitoria es que se trata de una red online de interacciones (Zysman y Kenney, 2018). En esta se produce el intercambio de información a través del uso de una computadora con conexión a internet (Varían, 2010; Haskel y Westlake, 2017).

En el contexto de la sociedad de la información, y desde una perspectiva clásica, la economía de la información está caracterizada por la información como un bien y su asociado que es la tecnología (Shapiro y Varían, 1998; Stiglitz 2000; Seidl, 2021). La información es el bien principal y es básicamente cualquier cosa que pueda ser digitalizada (Shapiro y Varían, 1998). Toda página *web*, datos, libros, entre muchos otros, constituyen bienes de información. Luego, existe una demanda que está dispuesta a pagar un precio por esta.

Si bien existe un costo inicial para producir información, hay un muy bajo costo para reproducirla lo que da lugar a la existencia de costos marginales tendientes a cero (Shapiro y Varían, 1998; Moazed y Johnson, 2016). Una implicancia directa que surge es que el precio, no viene determinado en base a los costos – en términos de la teoría clásica de formación de precios[11]–. Dado que cada consumidor tendrá un mayor o menor interés por este bien, la organización tiene la posibilidad de establecer un precio diferente para un mismo bien. Es decir, existirán precios diferenciales (Shapiro y Varían, 1998).

[11] Para la cual el mercado es un espacio abstracto en el que la demanda agregada y la oferta interactúan entre sí hasta producirse un ajuste que determina el precio (Coumot, 1838).

Además, debido al gran flujo de información que circula rápidamente por internet la organización puede contar con mayor conocimiento sobre el comportamiento de los usuarios. Esto le permite realizar un perfilamiento más acabado del individuo (Custers, 2013; Haskel y Westlake, 2017; Ghandy Jr., 2021) obteniendo información precisa sobre sus intereses o acciones. A partir de ello puede generar nuevas estrategias de acción a seguir, siendo posible gracias al poder extractivo y predictivo de las tecnologías de *Big Data* (Schmarzo, 2013; Constantiou y Kallinikos, 2015; Kolanovic y Krishnamachari, 2017: Culpepper y Thelen, 2019).

De este modo, el poder surge a partir de que las plataformas se convierten en agentes y mecanismo esenciales que conforman los mercados digitales modernos (Mazzucato, Entsminger, & Kattel, 2020). Pero particularmente se trata de un poder de plataforma (Culpepper y Thelen, 2019). Se origina a partir de la dependencia que las organizaciones generan para con los individuos en la medida que les brinda una comodidad - o facilidad- para transaccionar. En este nuevo entorno, los individuos son inducidos a ceder sus datos para poder satisfacer sus necesidades sin mediar contrato alguno (Zuboff, 2015, 2019).

Ahora bien, esta visión económica de mercado puede decirse que no resulta suficiente para estudiar el poder que ejercen las organizaciones sobre los individuos (Culpepper y Thelen, 2020). No solo se trata de un dominio en función de desarrollo (tamaño y escala). Sino que también es posible considerar el poder de cálculo que le brinda la tecnología a la organización convirtiéndola en un centro de cálculo[12] (Latour 1987; Callon y Muniesa, 2003, García Fronti y Herrera, 2021). De este modo, para concebir a este mercado de manera completa puede considerárselo como un ensamble de agentes o actores en una red de interacción digital.

Los autores Caliskan y Callon (2009) conciben al mercado como una red de ensamblajes sociotécnicos. En este sentido brindan una propuesta socio económica del mercado que resultaría adecuada para comprender la configuración de este nuevo entorno digital y como se produce el poder. El primer aspecto relevante es que existe una división entre los agentes o actores que participan. Por un lado, están los que crean la red de interacciones, y por el otro los elementos que se ensamblan. La capacidad de acción superior del centro de cálculo

[12] Esta concepción fue desarrolla en el apartado 1.2.1

(organizaciones) a través del uso intensivo de tecnología es lo que le permite adquirir una posición diferente y de superioridad en la red[13].

Desde esta perspectiva, es posible establecer tres elementos característicos que dan forma al mercado o red de ensamblajes sociotécnicos (Caliskan y Callon, 2010). El primero es que existe una circulación de bienes que en el entorno digital son los datos. Luego, se encuentran los dispositivos tecnológicos, así como el conocimiento científico, normas y convenciones sociales. Y finalmente, los agentes o actores (organizaciones e individuos) que establecen una disputa para realizar un acuerdo en la determinación de un precio para llevar a cabo un intercambio.

A partir de los elementos mencionados, surge la figura de agencia o '*agencements'* (Caliskan y Callon, 2010, García Fronti y Herrera, 2021). Su clave definitoria es la capacidad de acción de los agentes o actores que actúan en la red. Por un lado, están las organizaciones que poseen los dispositivos (tecnologías) y humanos (especializados) para llevar adelante la ejecución de cálculos obteniendo una posición de poder dominante. Por el otro, los individuos que si bien también son agentes de cálculo su capacidad de acción se encuentra limitada por el poder superior del anterior actor. Solo utilizan la red (usuarios) para satisfacer sus necesidades (Zuboff, 2019) a través de ceder sus datos.

De esta manera, el mercado en el contexto de la sociedad de la información se trata de un mercado electrónico (o digital o de plataforma) de ensambles sociotécnicos. El elemento característico es la tecnología que en principio actúa de intermediario entre los agentes. Ahora bien, esta no se trata solo de un instrumento productivo más, sino que es un dispositivo sociotécnico (Caliskan y Callon, 2010). Es gracias a la computarización o informatización de las acciones de los agentes en un entorno digital lo que permite la concreción del encuentro entre organizaciones e individuos. Pero, además, considerando la capacidad de agencia de los actores, es que resulta posible la ejecución de cálculos en base a datos por parte de las organizaciones. A partir de ello se produce la valorización que habilita el interés por el intercambio. De aquí que, sin la existencia de este valor, el mercado de datos deja de existir. Sin la información creada sería imposible encontrar los ensambles sociotécnicos reales que en él se producen para su generación.

[13] Es decir, construir un poder superior de acción en base a información como fue explicado en el apartado 1.1.3

1.3.2 La generación de valor en el mercado de datos

En la red sociotécnica que toma forma de mercado de datos, el valor se origina por la posibilidad de ejecutar cálculos (Callon y Muniesa, 2003; Çalışkan, y Callon, 2009, 2010; García Fronti y Herrera, 2021). Los agentes dotados de habilidades y al contar con los instrumentos tecnológicos necesarios logran generar valor. De esta manera, es la propia configuración algorítmica – esa secuencia de pasos lógicos como fue definido en el apartado 1.1.1.– de la red digital a través de la cual esos cálculos se materializan. Por esta razón, puede decirse que la red o plataforma es posible considerarla un "centro de cálculo" (Latour, 1987).

Si bien resulta cierto que en ocasiones existen precios diferenciales en esta nueva configuración en red, también es importante considerar que en ocasiones un precio se calcula en función de otro existente (Çalışkan, y Callon, 2010). Esta forma de cálculo incluye criterios sociales como por ejemplo la equidad. Un precio puede resultarle justo a un individuo, no por el precio en sí mismo sino por su forma de cálculo. Un caso donde es posible encontrar lo mencionado es cuando los precios se establecen por alguna norma de una agencia central. Por ejemplo, en el caso de Argentina, el programa de Precios Justos[14] creado por las Autoridades del Estado argentino. Estos precios no son cuestionados o recalculados por los ciudadanos, sino que son aceptados como neutrales, equitativos, justos ya que vienen determinados por una forma de cálculo normada.

De este modo, la generación de valor en esta red puede ser visto como un proceso de co-creación a partir de los ensamblajes producidos (Callon y Muniesa, 2003; Çalışkan, y Callon, 2009, 2010; García Fronti y Herrera, 2021). En este intervienen también cuestiones sociales. En el caso particular del valor público suele darse como un proceso colaborativo (co-creado) o de acuerdo entre los actores intervinientes (Bryson et al., 2015). Si se piensa en el ejemplo de Precios Justos anteriormente mencionado, surge del acuerdo entre productores, los ciudadanos y el Estado. De aquí que, en la concepción moderna de la gestión pública, el Estado tiene un rol especial como garante de los valores públicos (Bryson et al., 2015). Esto porque contribuye a generar el bienestar colectivo. Es entonces que el valor público creado en el escenario planteado se torna de vital importancia.

Pero al mismo tiempo, sucede que, en este proceso de creación de valor, se genera una dependencia de los ciudadanos con la red. Esto le otorga mayor poder de control sobre los

[14] Acuerdo de precios para ciertos productos. Más información en https://www.argentina.gob.ar/economia/comercio/preciosjustos

ciudadanos al Estado (Zuboff, 2015; Gandy Jr, 2021). Se debe a que los datos obtenidos de los ciudadanos son utilizados como medios para la modificación de conductas humanas. Esto ocasiona que las organizaciones logren concentrar mayor poder de control[15]. Pero, además, la plataforma digital de gobierno se convierte en un recurso central para que la extracción de datos se lleve a cabo (Mazzucato, Entsminger, & Kattel, 2020). En base a estas perspectivas, el mercado de datos también resulta en un mercado de datos conductuales.

En este sentido, Zuboff, (2015) plantea la existencia de una relación subjetiva entre la red y los individuos que carece de consentimiento explícito. Para convertir esta subjetividad en resoluciones de objetivos concretos (políticas públicas), las organizaciones públicas del estado utilizan los dispositivos de cálculo. A su vez, esto genera un ciclo de retroalimentación donde ya no existe contrato alguno entre partes sino más bien un poder de vigilancia (Zuboff, 2015). Y esto es sostenido en el tiempo por la existencia de una asimetría de información por medio de la cual la organización logra obtener más conocimiento sobre el ciudadano.

1.3.3 La asimetría de información como parte constitutiva del poder de acción de las organizaciones

La asimetría de información que se produce en el mercado de datos confluye en una asimetría de poder en manos de la organización. En el caso de las organizaciones públicas es aprovechado para intervenir con políticas públicas de control. Les permite realizar una gobernanza anticipatoria reduciendo la complejidad social futura a un problema del presente (Guston, 2014; Aykut, Demortain, & Benbouzid, 2019). En cambio, en las organizaciones de carácter privado, es utilizado para maximizar sus beneficios. Ahora bien, las causas que dan origen son diversas. Pero consideradas en conjunto toman la forma de asimetría de conocimiento o información. Para lograr su comprensión en esta sección se desarrolla una explicación.

El concepto de asimetría de información surge en el contexto de mercado (clásico) y fue introducido por Arkelof[16] en 1970. En su trabajo "*The Market for "Lemons": Quality Uncertainty and the Market Mechanism*" teoriza acerca de que si se produce un desequilibrio en la cantidad de información que poseen oferentes y demandantes no se producirá la

[15] Como fue explicado en el apartado 1.1.3
[16] Economista americano nacido en 1940. Actualmente, profesor en Georgetown University y profesor de economía emérito en University of California.

transacción resultando en una falla del mercado (Aoun Barakat & Sayegh, 2021). El hecho de que el oferente tenga más información produce una incertidumbre para el demandante acerca del producto a intercambiar. En este mismo sentido, pero en el contexto del mercado de datos definido en el apartado 1.3.1, son las organizaciones las que basan sus decisiones en datos ya que poseen más información que los ciudadanos.

A través del uso intensivo de la tecnología del *Big Data*, la organización posee una mayor capacidad de cálculo que le genera un mayor poder predictivo de información. Pero el proceso realizado es omnipresente en la red de interacciones por la falta de transparencia de los algoritmos utilizados. Los individuos no pueden conocer cómo funcionan, pero tampoco saben que datos son los que utilizan (Zuboff, 2019; Cutolo & Kenny, 2021; Ghandy Jr., 2021). Podría decirse entonces que para el individuo se genera una incertidumbre en el sentido definido en el párrafo anterior y mayor poder de acción para la organización.

Pero, además, los datos, en particular los personales, son utilizados para modificar las conductas humanas (Zuboff, 2015; Ghandy Jr., 2021), limitando la capacidad de acción del individuo en la red o ejerciendo un control sobre la misma (Zuboff, 2019; Cutolo & Kenny, 2021). En este sentido, la organización posee un poder de control sustentado en una asimetría de conocimiento que para ella elimina la incertidumbre. En el caso de las organizaciones privadas, esto es monetizado en el mercado permitiéndole obtener un mayor rédito (Aboody & Lev, 2000). En cambio, las organizaciones públicas lo utilizan para intervenir con políticas públicas peor al mismo tiempo ejerciendo un poder de vigilancia o de control (Zuboff, 2019; Mazzucato, Entsminger y Kattel, 2020; Ghandy Jr., 2021).

Concretamente, la asimetría de información generada en el mercado de datos refiere a las diferencias sustanciales existentes. La información que son capaces de producir las organizaciones mediante la realización de cálculos en base a los datos de los ciudadanos es muy superior a la que se encuentra disponible para estos (van de Waerdt, 2020). A partir de ello puede colocarse al individuo en una situación de vulnerabilidad ya que limita su capacidad de acción como se ha mencionado anteriormente. Pero también porque existe el riesgo de que alguno de sus derechos pueda verse afectado. Si bien cierta parte de la población acepta que se produzca la recopilación de sus datos personales a cambio de satisfacer una necesidad, otra parte considera que trae aparejado consecuencias. En particular, sobre sus vidas privadas en el sentido de que la utilización de la tecnología resulta cuanto menos invasiva generándoles preocupación (Gandy Jr, 2021).

Con ello, queda en evidencia el poder superior de acción que poseen las organizaciones. Se sustenta en una asimetría de información en el contexto global de la sociedad de la información en red y del mercado de datos en particular. Si bien puede ser utilizada con fines bien intencionados, también existe el riesgo de que suceda lo contrario. Esto invita a exponer la responsabilidad que conlleva el contar con una capacidad de cálculo tan poderosa, particularmente en el caso de las organizaciones públicas del estado. Particularmente, la protección de datos personales se convierte en un punto central para contribuir a la continuidad de un sistema de toma de decisiones democráticamente participativo. Este riesgo asociado a privacidad de datos será abordado con mayor detalle en el capítulo 2.

No obstante, surge la necesidad de contar con pautas claras para desarrollar una gobernanza de datos eficiente y eficaz en la protección de datos en contextos organizacionales. El conocimiento científico existente puede ser utilizado también para ello. Al contar con métodos cuantitativos desarrollados puede contribuir en este sentido como ser el caso de la Privacidad Diferencial. De este modo los resultados obtenidos del uso de tecnologías y conocimiento ya no solo deberían ser empleados como mecanismo de ejercicio de poder o control. Si no también como un elemento que contribuya a reducir los riesgos asociados de forma tal de contribuir a brindar garantías de privacidad para los individuos, entre otros.

Conclusión del capítulo

En el presente capítulo, se ha presentado como en el contexto de la sociedad de la información las organizaciones públicas del Estado construyen un poder superior de acción frente a los individuos. Se ha analizado como crean valor público mediante la obtención de los datos de los ciudadanos, causando una asimetría de poder. La utilización de tecnologías de *Big Data* para la aplicación de conocimiento a los datos les permite convertirlos en un medio para la modificación de sus acciones posibles en un escenario de mercado. En este sentido, se trata de un poder para el ejercicio del control. En ocasiones puede resultar necesario como en un caso de salud pública en un contexto de pandemia. En otras situaciones, podría incurrirse en la violación de derechos como ser el de privacidad.

Se ha propuesto que la sociedad de la información es una red que viene configurada por el accionar de los actores visible a través de los rastros que dejan al interactuar entre sí en un entorno digital. En esta, el procesamiento de datos es omnipresente para los ciudadanos, pero creado y controlado por las organizaciones. En este sentido es posible colocar al individuo en el centro de circulación de la información y notar como su capacidad de acción es limitada

por un poder de acción superior que poseen las organizaciones. Este poder es construido en base a información generada mediante recursos tecnológicos.

A partir de ello, es posible decir que el tipo de organización que se configura es un centro de cálculo en el cual las tecnologías de *Big Data* juegan un rol central. Al ser combinadas con el conocimiento científico y una gestión eficiente de datos, les permite convertir los escenarios futuros inciertos en problemas del presente. Particularmente, el Estado puede intervenir con la ejecución de políticas públicas para ejercer un control, una gobernanza anticipatoria. De este modo, las organizaciones públicas del estado logran posicionarse con un poder de acción superior frente al ciudadano.

A su vez, la configuración de una organización como un centro de cálculo es lo que le permite interactuar como un actor de intercambio en el mercado de datos. Pero este mercado ya no es entendido en el sentido económico clásico, sino que surge de un proceso de co-constitución. Los agentes sociotécnicos dotados de capacidad de acción se ensamblan para poder interactuar en un espacio digital. Esto es posible dada una configuración algorítmica denominada plataforma que es creada por las organizaciones. En el contexto estatal, tal configuración se denomina gobierno digital en donde las organizaciones públicas encuentra la potenciación de su capacidad de acción. Ello porque logran extraer los datos pertenecientes a los individuos y ejecutar cálculos.

A partir de tal configuración del mercado es posible hablar de un mercado de datos. El valor público se origina por la posibilidad de ejecutar cálculos por parte de los agentes dotados de esta habilidad. Junto con los dispositivos tecnológicos que poseen es entonces que las organizaciones pueden materializar el valor. Pero al mismo tiempo se produce una asimetría de información. No se trata de una simple asimetría negativa en términos clásicos. La cantidad de información generada en base a los datos recolectados les permite a las organizaciones públicas conocer más sobre los individuos, incluso realizando un perfil acabado de los mismos. Al ser utilizada como gobernanza anticipatoria para modificar los escenarios de acciones posibles de los ciudadanos, logran construir un poder de control superior. En ocasiones puede resultar necesario para garantizar el bienestar general. Pero en otras puede utilizarse con fines excesivos o mal intencionados omitiendo los derechos que poseen los ciudadanos.

Frente al anterior escenario planteado, el individuo queda en una situación de vulnerabilidad. Esto porque surge el riesgo de que alguno de sus derechos pueda verse afectado.

Particularmente, en algunos casos puede resultar necesario la utilización de datos personales para proteger a la población en su conjunto. Pero también expone como el uso de información implica una gran responsabilidad por parte de las organizaciones públicas. Esto porque incluso podría ponerse en riesgo la privacidad de los individuos o el derecho que poseen a controlar que es lo que se hace con sus datos. Por esta razón, sería conveniente considerar la utilización de metodologías para dar garantías de protección de datos personales.

A diferencia de una ley o normativa, la utilización de metodologías cuantitativas para protección de datos posibilitará reducir los posibles riesgos asociados de manera preventiva antes que reactiva. Para lograrlo deberá incorporase desde el diseño del proceso de datos en el contexto organizacional dentro de un marco de responsabilidad. De este modo se podría brindar cierta garantía de privacidad contribuyendo a un sistema democráticamente más justo. En el capítulo a continuación se aborda la problemática de privacidad en base a datos para poder alcanzar una mayor profundización al respecto.

Capítulo 2: La privacidad en la sociedad de la información

Introducción

En el contexto de la sociedad de la información definido en el capítulo previo, donde los actores se interrelacionan a través de una red digital, surgen diferentes cuestionamientos entorno a la privacidad. Específicamente, en torno a la privacidad de datos personales. Esto se debe a que se encuentra potencialmente amenazada por la omnipresencia del proceso de datos que llevan adelante las organizaciones. En Argentina, la normativa o legislación vigente pareciera no ser garantía suficiente. Diferentes hechos de conocimiento público como ser el robo de datos personales de jubilados de la base de datos de PAMI (Instituto Nacional de Servicios Sociales para Jubilados y Pensionados) en agosto de 2023[17], la violación de acceso a servidores del poder judicial de la provincia de Chaco en enero de 2022[18] o la denuncia presentada por el uso irrestricto de datos biométricos en un organismo del Gobierno de la Ciudad de Buenos Aires en 2023[19], entre otros, exponen el riesgo al cual quedan expuestos los ciudadanos cuando sus datos son vulnerados.

[17] https://www.cronista.com/economia-politica/jubilados-pami-se-filtraron-los-datos-de-los-beneficiarios-que-cuidados-tomar/
[18] https://www.iproup.com/innovacion/28826-hackean-los-servidores-del-poder-judicial-de-chaco
[19] https://www.ambito.com/politica/denuncian-penalmente-la-ciudad-el-uso-datos-biometricos-justificacion-racional-n5705668

Además, con la intensificación del uso de dispositivos tecnológicos, comenzaron a surgir discusiones desde diferentes ámbitos sobre la privacidad de datos personales (Smith, Dinev, & Xu, 2011; Friedman, Kahn, Borning & Huldtgren, 2013; Mortier et.al., 2016, Ghandy Jr., 2021). En muchas ocasiones posterior a que un incidente haya expuesto el problema. Particularmente, esto se ha incrementado desde la última pandemia que comenzó en el año 2019 alcanzando al mundo entero. A partir de este suceso comienza a resonar fuertemente el debate acerca de cómo se ve afectado el derecho a privacidad a partir del uso de datos personales por parte de las organizaciones.

Como consecuencia ha surgido diferente regulación en materia de privacidad de datos. Entre estas, surgen dos que presentan enfoques muy diferentes. Por un lado, la norteamericana que posiciona al individuo como un consumidor (Varían & Berkeley, 1996; Smith, Dinev, & Xu, 2011; Arner, Castellano y Selga, 2022) y utiliza la herramienta legislativa como soporte para asignarle la responsabilidad de control de su privacidad. Por el otro, la regulación europea que reconoce a la privacidad como un derecho humano y busca contrarrestar la concentración en manos privadas (RGPD, 2016; AEPD, 2019, 2020; Arner, Castellano y Selga, 2022). Entre medio de estas, se encuentra la legislación argentina con su Ley 25.326 y su marco constitucional.

En términos generales, la legislación actúa una vez que un hecho ya sucedió. En este sentido, se la puede interpretar como reactiva antes que preventiva. Si bien, resulta necesaria para vivir en sociedad como medio para reparar un daño o establecer un castigo, presenta una limitación. De manera genérica, podría decirse que no resulta suficiente para definir normas y procedimientos para mitigar riesgos. En el contexto particular de datos, la responsabilidad involucrada que conlleva su procesamiento y uso de parte de las organizaciones no es alcanzada por la normativa (Rubinstein, 2012). A partir de ello se abre la posibilidad de plantear él porque es necesario una gestión responsable de datos personales en organizaciones. Desde una perspectiva de responsabilidad en torno al respeto del derecho humano y de uso de tecnología (Cavoukian, 2011) sobre datos personales, esto podrá ser sintetizado en la idea de privacidad desde el diseño y por defecto.

En este contexto, y considerando el marco expuesto en el capítulo anterior, surge el siguiente interrogante: ¿cómo es posible definir a la privacidad de datos personales en el contexto de la sociedad de la información desde una perspectiva de gestión responsable? Para responder esta pregunta el objetivo a desarrollar en el presente capítulo consiste en elaborar una

definición de la privacidad de datos personales. Partiendo de una concepción de la privacidad como un derecho humano, luego se la relacionará con la protección de datos personales desde una perspectiva operativa. De este modo se podrá construir una definición pero que considere una perspectiva ética, la normativa vigente, así como las implicancias en términos de responsabilidad de gestión por parte de las organizaciones.

A fin de elaborar una definición de la privacidad de datos personales, el presente capítulo se divide en tres secciones. En la primera sección se desarrolla la conceptualización de la privacidad desde una perspectiva operativa en base a datos. Ante la existencia de un procesamiento omnipresente de datos, se coloca al individuo en el centro de este para plantear un abordaje desde una visión ética responsable del derecho humano. Sobre esta base, se amplía el concepto de privacidad a privacidad de datos personales considerando el contexto de la sociedad de la información desarrollado en el capítulo 1. Finalmente se elabora una definición de la privacidad de datos personales.

La segunda sección, se centra en la regulación vigente a partir de tomar como referencia a Estados Unidos de América y la Comunidad Europea como posturas dominantes en el mundo occidental. Luego se hace referencia al caso de Argentina. Para ello se comienza por presentar las perspectivas de los dos primeros casos mencionados y luego se aborda el caso argentino en particular. Finalmente se realiza una comparación entre los tres casos. De esta surgirá que en el caso argentino no existe un marco estandarizado para la protección de datos personales en organizaciones. De aquí que se considerará necesario como medio para brindar garantías de privacidad en los términos de la definición propuesta.

En la última sección, se desarrolla acerca de la propuesta de concebir a la privacidad de datos personales desde la idea de la privacidad operativa. Esta implica abordarla desde un enfoque de gestión de datos organizacional para el diseño de sistemas que la contemplen y se asignen responsabilidades por defecto. Para ello en primer lugar, se desarrolla sobre la concepción ética del uso de tecnología para el procesamiento de datos. Luego, se continúa con el abordaje del concepto privacidad de datos personales por defecto para exponer la relevancia de la responsabilidad de las organizaciones a la hora de procesar datos. Finalmente se desarrolla el concepto de privacidad de datos personales desde el diseño bajo una perspectiva ética operativa. En concreto implicará definir los principios que deberán tenerse en cuenta para gestionarla bajo un marco de responsabilidad. De aquí surgirá la necesidad de establecer

una metodología cuantitativa para brindar garantía de privacidad en términos de una gestión de datos responsable por parte de las organizaciones.

2.1 La privacidad en base a datos desde un enfoque ético y de derecho humano

Existen diversas perspectivas para el abordaje del concepto de privacidad al tratarse de un valor social. Ha sido estudiada por diferentes disciplinas a través de los años aportando diferentes visiones, alrededor de una concepción ética. No obstante, existe una primera distinción generalizada entre lo que se refiere a la privacidad física y a la privacidad de identificación individual. Mientras la primera refiere al acceso físico sobre un individuo y/o su entorno; la segunda refiere al uso de información para la identificación de un individuo (Smith, Dinev, & Xu, 2011). Esta distinción resulta fundamental para poder abordar a la privacidad dentro del contexto de la sociedad de la información como se ha planteado desde el inicio del presente trabajo.

Considerando lo desarrollado en el capítulo 1, es posible colocar al individuo en el centro del procesamiento de datos para reflexionar sobre el impacto que la tecnología aplicada a datos personales causa entorno a su privacidad. En este sentido, el individuo puede ser considerado un actor en el mercado de datos que transacciona datos pero que también posee derechos. A partir de aquí se abre la posibilidad de pensar en la existencia de un contrato social (tácito), una confianza depositada en las organizaciones por parte de los individuos. A partir de ello, y en términos económicos, es posible el intercambio. Ahora bien, para garantizar una igualdad de oportunidades en base al respeto de derechos resultaría adecuado establecer un marco de responsabilidad sobre el uso de datos. Esto porque la asimetría de información generada posiciona al individuo en una situación de desventaja frente al poder de las organizaciones.

De este modo, la existencia de una red digital de interacción (mercado de datos) posee un actor con mayor poder de acción que son las organizaciones. Estas crean un proceso de datos que se encuentra omnipresente en la red y los individuos se encuentran en una situación de desventaja frente a los riesgos que se suscitan. Bajo esta situación, en este trabajo la privacidad será abordada desde una perspectiva de responsabilidad y sustentada en un enfoque normativo. El aspecto normativo implica una concepción basada en el respeto de derechos humanos donde el Estado es el garante para lo cual resulta adecuado que actúe con el ejemplo. A su vez, el enfoque propuesto será llevado al contexto de un marco operativo a la hora de pensar en la responsabilidad involucrada en las acciones de los actores.

Históricamente, la privacidad como un derecho de los individuos encuentra su origen al haber sido declarada en 1948 como un derecho humano universal por la Organización de Naciones Unidas[20] (ONU). En el artículo 12 de la Declaración Universal de Derechos Humanos se establece que "Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques." (pág. 26). Este artículo sentó la base para que los diferentes países, en particular del mundo occidental, comenzaran a regular a la privacidad de diferentes maneras según el sistema legal vigente y cultural. Pero, principalmente se concebía a la privacidad como la privacidad física.

Con la revolución tecnológica desde finales del siglo XX, frente a la posibilidad de obtener conocimiento en base a datos es que se comienza a poner mayor énfasis en la problemática de privacidad. Pero a diferencia del pasado, ahora se lo hace desde la perspectiva de la identificación de individuos mediante el uso de datos personales. En Argentina, no será hasta el año 2000 cuando se sanciona la Ley 25.326[21] de Protección de Datos Personales en que la identificación de individuos a partir del uso de datos personales quede efectivamente regulada o normada. A partir de ello se desprende una consecuencia directa que es que el individuo tiene el derecho a controlar quien puede acceder y quien no a las cuestiones de índole personal también en base a datos.

Bajo esta perspectiva, la privacidad es definida como el control que un individuo tiene sobre el acceso que otros podrían tener a cuestiones personales como ser características propias que decide mantener como privadas (Gandy Jr, 2021). También, las posibles de obtener a partir del procesamiento de datos personales. Cuando este control se pierde, surge entonces una preocupación en los individuos, así como diferentes cuestionamientos. Particularmente, en la red digital de interacción entre ciudadanos y organizaciones, comienza a cuestionarse cuál es el nivel de acceso a información de índole personal que se produce. Pero se da un paso más, al preguntarse qué implicancia tiene el procesamiento de datos en términos de pérdida de control de la privacidad para el individuo (Gandy Jr, 2021). En este sentido, la

[20] La ONU es una organización para "el compromiso del mundo para la promoción y protección del conjunto total de derechos humanos y libertades establecidos en la Declaración Universal de Derechos Humanos" la cual se conforma por países miembros entre los cuales se encuentra Argentina.
[21] Ley 25.326 https://servicios.infoleg.gob.ar/infolegInternet/anexos/60000-64999/64790/texact.htm

privacidad es abordada desde la perspectiva del riesgo de identificación de individuos a partir del uso de tecnología aplicada a datos personales, es decir, en un sentido operativo.

La discusión sobre este proceder alrededor del mundo entero ha crecido en la medida que se produce más y más recopilación de datos. Los nuevos desarrollos tecnológicos han contribuido en un espacio altamente digitalizado. Sin ir muy atrás en el tiempo, con el comienzo de la pandemia de COVID-19 a nivel mundial en el año 2019 se ha utilizado de manera intensiva diferentes dispositivos tecnológicos para la recopilación de información personal por parte de los Estados. Bajo la justificación de tratarse de un tema de salud pública, comenzaron a surgir discusiones acerca de cómo se veía afectado el derecho a privacidad de los ciudadanos. Es decir, comienza a tomar mayor relevancia y preocupación la sensibilidad de la información recopilada (Gandy Jr, 2021).

Si bien, en la sociedad moderna bajo un sistema democrático, el Estado es quien tiene el rol de garante de los valores públicos y derechos individuales, este también se conforma de organizaciones que procesan datos. También generan una asimetría de información como se explicó en el apartado 1.3.3, a partir de utilizar tecnologías de *Big Data* para la aplicación de conocimiento a los datos para la obtención de información. Su utilización les brinda el poder de convertirlos en medios para la modificación del escenario de acciones posibles de los ciudadanos colocándolos en una situación de desventaja. En este sentido, puede interpretarse que se trata de un poder de control, que en ocasiones es necesario para contribuir al bienestar social– como puede ser un caso de salud pública en un contexto de pandemia-. Pero también, puede incurrir en la violación de la privacidad ante lo cual pareciera que la normativa existente si bien es necesaria puede resultar insuficiente. A pesar de su existencia en el caso argentino han ocurrido diferentes hechos de conocimiento público como los mencionados anteriormente sobre accesos no autorizados a servidores de organismos públicos para el robo de datos de los individuos. Incluso, también la publicación de identificadores personales por parte de alguna organización pública del estado[22].

En el contexto de la sociedad de la información surge una nueva perspectiva operativa tecnológica operativa para conceptualizar a la privacidad como la privacidad de datos personales. Esto porque, como se explicó en el apartado 1.3, existe un mercado de datos

[22] Por ejemplo, la AGIP (Administración Gubernamental de Ingresos Públicos) de la Ciudad de Buenos Aires tiene publicado un listado de CUIT (Clave Uniforma de Identificación Tributaria) que permite identificar unívocamente a los ciudadanos argentinos, en su portal de datos abiertos: https://data.buenosaires.gob.ar/dataset/alto-riesgo-fiscal

donde se produce una interacción entre organizaciones, tecnología e individuos. Pero también donde se lleva a cabo un procesamiento de datos omnipresente que solo las organizaciones tienen el poder de crearlo y controlarlo. Frente a ello queda expuesta la vulnerabilidad bajo la que se encuentra el individuo al perder la posibilidad de controlar quien y como se accede a su información personal.

2.1.1 Definición de la privacidad como la protección de datos personales

En el contexto de la sociedad de la información, al considerar la privacidad de datos personales, permite incorporar nuevas dimensiones de análisis sobre la privacidad (Smith, Dinev, & Xu, 2011). No solo se trata de la privacidad física, sino que también es posible incurrir en su violación debido al potencial de procesamiento que la tecnología brinda cuando es utilizada para aplicar conocimiento a datos personales. A partir de ello se abre la posibilidad de evaluar el riesgo asociado en términos de identificación directa o indirecta de individuos en base a datos personales.

Los datos personales refieren a un atributo o conjunto de atributos que permiten realizar un perfilamiento del individuo (Custers, 2013; Haskel y Westlake, 2017; Ghandy Jr., 2021). Esta acción deriva en una identificación unívoca del ciudadano. Por ejemplo, a través del nombre y apellido, número de documento, CUIT. Pero también los identificadores únicos de los dispositivos móviles, datos de geolocalización, así como datos biométricos, también conforman parte de estos. Por otro lado, los identificadores indirectos (como, por ejemplo, género, código postal, entre otros) son aquellos que recombinados permiten realizar una identificación indirecta del individuo. Latanya Sweeney (2000) ha demostrado cómo es posible re identificar individuos en una base de datos a partir de combinar código postal, género y fecha de nacimiento. Particularmente sobre los datos del censo de 1990 de Estado Unidos de América, logró identificar unívocamente al 87% de la población de ese país a partir de esos tres identificadores indirectos.

A su vez, todos esos datos personales analizados en conjunto permiten conocer el comportamiento de un individuo. Por ello resultan muy atractivos para ser recolectados por las organizaciones. Que estas utilicen los datos de los individuos permite ampliar la discusión sobre la privacidad a partir de dos dimensionalidades específicas: el uso indebido de datos personales y la recombinación de datos personales. Como sostiene Gandy Jr. (2021), el uso indebido de datos personales implica tomar dimensión de los aspectos personales a diferencia de los aspectos públicos del individuo. De este modo será posible

considerar el potencial daño para estos que podría resultar si se hace un uso no autorizado o un uso incorrecto de sus datos.

Además, la recombinación de datos personales ha dado lugar a cuestionar acerca de la posibilidad de hacer recombinación de información personal proveniente de diferentes fuentes. A partir de ello preguntarse cómo se ve afectada la privacidad del individuo (Gandy Jr., 2021). Particularmente, en las esferas de las organizaciones del Estado es donde más se conoce (públicamente) que esto es realizado por lo cual se pone en duda este punto con mayor frecuencia. La incorporación de datos personales en combinación con otros puede provocar la creación de información sin el consentimiento de sus titulares, aunque la intención de su uso no sea mal intencionado. Si bien, en el caso particular de las organizaciones públicas, podría decirse que en la mayoría de las ocasiones son utilizados para favorecer a los ciudadanos a través de políticas públicas. Pero también pueden ser manipulados en su perjuicio, pudiendo incurrir en la violación de su privacidad (AEPD, 2019, 2020).

Considerando el contexto organizacional basado en datos, los riesgos asociados a privacidad pueden derivarse de cuatro acciones específicas: reciclar, reutilizar, recombinar y reanalizar (Steinmann, Matei y Collmann, 2016). La conjunción de estas cuatro acciones puede llevar a la generación de nueva información que excede al objetivo primario por el cual el individuo cedió sus datos, como fue explicado en el apartado 1.2.3. En este punto puede surgir un conflicto ético ya que el accionar de las organizaciones puede ser sin consentimiento. Por esta razón, el riesgo en términos de privacidad aumenta y resulta necesario una gobernanza responsable de datos.

En este contexto, surgen cuatro términos claves asociados a la privacidad de datos personales que permiten sentar la base para construir una definición de esta. El primero de ellos es el "anonimato". Refiere a cuando un individuo decide no hacer visible (o ceder) sus identificadores como por ejemplo sus datos personales (Smith, Dinev, & Xu, 2011). Si esto es realizado, entonces se imposibilita para una organización poder correlacionar información con el individuo en el contexto digital. Pero a su vez, el individuo no podrá construir una interacción con esta. Si bien se presenta como una acción de control posible para el individuo, podría decirse que por sí solo el anonimato no hace al concepto de privacidad de datos en el marco de responsabilidad.

Un segundo término clave asociado a la privacidad es "el secreto" en el sentido de que el individuo retiene información (oculta) para que no sea posible identificarlo principalmente en entornos digitales (Smith, Dinev, & Xu, 2011). Por ejemplo, mediante bloqueo de acceso a su IP (dirección única en internet). Ahora bien, el ocultar información tampoco hace a la privacidad en sí. Si no es posible construir un comportamiento a partir de la información personal, no será posible determinar si es moral o éticamente aceptable el comportamiento del individuo en términos sociales.

Un tercer término es la "confidencialidad". En el contexto de la sociedad de la información, este término refiere a la cantidad de información personal que es cedida por el individuo. El individuo solo otorga los datos precisos o básicamente necesarios a una organización en específico (Smith, Dinev, & Xu, 2011). Podría decirse que la confidencialidad es la posibilidad para el individuo de otorgar información bajo ciertas condiciones a una organización. Pero el derecho a la privacidad refiere a la posibilidad de poder tener conocimiento (y control) sobre quien, y como se accede a su información individual. En este sentido, tampoco este término por sí solo hace a la definición de privacidad de datos personales.

Por último, surge el término de "seguridad" que principalmente en un contexto de red digital se asocia directamente a la perspectiva tecnológica. Por este se hace referencia a que es necesario brindar seguridad informática sobre los datos lo que permitirá construir un accionar técnico éticamente responsable en una organización. Por ejemplo, establecer los mecanismos de seguridad informática necesarios para evitar un hackeo de una base de datos o restringir el acceso a esta. Pero, aun así, ello no evita la reutilización, recombinación o posterior análisis de los datos (Smith, Dinev, & Xu, 2011). En este sentido, la dimensionalidad del impacto causado en la privacidad del individuo por el procesamiento de datos a través de tecnologías de *Big Data* no sería alcanzado por este término.

Ahora bien, tomando como base la perspectiva ética que da sustento a la determinación de la privacidad como un derecho humano y considerando todo lo expuesto previamente, los cuatro términos claves (anonimato, secreto, confidencialidad y seguridad) en conjunto resultan en partes constitutivas de la privacidad de datos personales. Esto porque cada uno de ellos implica una cierta responsabilidad para quien lleve adelante un procesamiento de este tipo de datos para contribuir en brindar cierta garantía de privacidad a sus titulares. Pero, además, implicará tener en cuenta a la privacidad desde la fase de diseño de los procesos de

datos de forma tal que se apliquen los métodos necesarios para otorgar garantías de protección de datos personales.

De este modo, en este trabajo, se define a la privacidad de datos personales como la posibilidad de que un individuo generador de datos al interactuar con una organización en un contexto de red digital pueda obtener garantías suficientes de que sus datos están siendo debidamente protegidos dentro de un marco de responsabilidad. Se considera que de esta definición surgen dos implicancias concretas. Por un lado, se requerirá que la organización considere a la privacidad desde el diseño y por defecto para el procesamiento de datos en un marco de responsabilidad. Por el otro, que el individuo tenga la posibilidad de ejercer su derecho a decidir si acepta o no que se continúe con el procesamiento de sus datos en base a las garantías otorgadas.

Además, se considera que dicha definición de privacidad de datos personales respeta la Ley 25.326 de Protección de Datos Personales en el caso de Argentina que será abordada con mayor especificidad en la sección siguiente. Pero, también da la posibilidad de incluirla dentro de ciertos lineamentos generales vigentes desde el año 2013 propuestos por la Organización para la Cooperación y el Desarrollo Económicos (OCDE)[23] de la cual forma parte la República Argentina. Estos lineamientos abarcan los cuatros términos claves definidos con anterioridad y constituyen un marco general para definir políticas claras para el resguardo de datos personales en contextos organizacionales.

Los lineamientos generales de la OCDE (2013) postulan que se debe establecer límites claros para la obtención de los datos personales, así como determinar la relevancia de estos para el uso previsto. En cuanto a su recolección, establecen la necesidad de definir con claridad el uso que se dará a los datos antes de solicitárselos a sus titulares. También se propone abstenerse de utilizar los datos para usos distintos al determinado originalmente sin el consentimiento de las personas afectadas. Se deberán proteger contra el acceso ilícito o piratería. Así mismo se deberá garantizar que las personas cuyos datos se han recolectado tengan acceso a los mismos y puedan solicitar modificaciones o su eliminación definitiva.

De esta manera, la gestión en torno de la privacidad de datos personales en contextos organizacionales se convierte en un punto central para contribuir en la construcción de

[23] Personal Data Protection at the OECD [En línea]: https://www.oecd.org/general/data-protection.htm#:~:text=The%20OECD's%20data%20protection%20rules&text=The%20rules%20require%20that%20personal,for%20no%20longer%20than%20necessary.

garantías de privacidad. Ello requerirá que las organizaciones actúen bajo consideraciones éticas de privacidad, en términos de la ley. Pero también asumiendo las responsabilidades que se derivan de los lineamientos generales considerados en este apartado.

2.2 Principales enfoques regulatorios de la privacidad de datos personales

Con la intensificación del uso de tecnología para el procesamiento de datos personales desde finales del siglo XXI, surgen consecuencias que afectan la privacidad de los individuos. A partir de ello, los diferentes países del mundo sustentados en una perspectiva ética del deber ser han creado leyes y/o regulaciones. Cada uno en el marco de su propio sistema de gobierno para tratar de controlar los efectos de este nuevo fenómeno.

Según datos publicados por la Organización de Naciones Unidas (ONU) a noviembre 2023, 137 países de 194 en el mundo – en base a los reconocidos por esta organización– han promulgado leyes buscando dar garantía de protección de datos y privacidad[24]. Que el 71% de los países ya cuenten con una legislación establecida en el año 2023 sobre protección de datos personales, por un lado, permite dimensionar la relevancia de la temática. Por el otro, habilita a preguntarse si contar solo con esta resulta suficiente. Para poder alcanzar una respuesta a continuación se desarrolla sobre el enfoque sobre privacidad de datos personales que los países utilizan.

A los fines de simplificar este análisis, se toma como referencia al enfoque de los Estados Unidos de América y de la Comunidad Europea, ya que se han convertido en los principales referentes para la mayoría de los países en Latinoamérica y particularmente para Argentina. Según Zwick y Dholakia (1999), el modelo propuesto en el primer caso es el de autorregulación y en el segundo caso se trata de una regulación. El primer modelo considera a la privacidad como una mercancía bajo el paragua del libre mercado. Los datos son vistos como bien para favorecer el desarrollo de la economía ya que son valorables e intercambiables (Arner, Castellano y Selga, 2022).

En cambio, en el modelo de regulación la privacidad es considerada como un derecho humano por lo que resulta necesario brindar garantías al individuo. Los datos son

[24] Data Protection and Privacy Legislation Worldwide [En línea]: https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

considerados un derecho individual. Al ser generados por el individuo se considera que se debe contrarrestar la concentración en manos privadas (Arner, Castellano y Selga, 2022).

A continuación, se específica sobre cada modelo para luego abordar el caso argentino. Esto brindará la posibilidad de realizar una comparación entre los elementos claves que poseen. Con ello se podrá comprender si la normativa resulta suficiente o si es necesario un enfoque preventivo como el propuesto en la definición de privacidad realizada.

2.2.1 El enfoque norteamericano

El enfoque norteamericano sobre la privacidad de datos parte de su comodificación sustentada en la idea dominante de libre mercado. Es considerada en base a los principios económicos de beneficio y costo en una relación de intercambio de mercado (Varían & Berkeley, 1996; Smith, Dinev, & Xu, 2011). En este sentido es propuesta como una falla de mercado (o externalidad negativa) como fue expuesto en el apartado 1.3.3. Por esta razón, es al propio mercado a quien se le solicita que resuelva esta falla. De aquí que el modelo propuesto es el de una autorregulación.

La perspectiva de mercado encuentra en la privacidad un componente económico. Por el lado de la demanda, se argumenta que el individuo proporciona voluntariamente sus datos en un contexto digital para lograr un intercambio en tanto agente del mercado (Smith, Dinev, & Xu, 2011; Arner, Castellano y Selga, 2022). Al concebir al individuo como un consumidor, se considera que es el mismo quien utiliza a la privacidad como una mercancía que puede intercambiar a cambio de algún beneficio. A su vez, el costo que asume es el de recibir información constantemente sobre diferentes bienes o servicios. Por ejemplo, a través de continua publicidad en las redes sociales o emails, entre otros (Varían & Berkeley, 1996).

Del lado de la oferta, el costo en que se incurre es en el envío de la oferta del producto o servicio. Este podría reducirse si el oferente conociera más acerca del potencial consumidor para determinar si le es conveniente enviarle información o no (Varían & Berkeley, 1996; Zwick y Dholakia, 1999). En este sentido, se habla de falla de mercado debido al costo en que incurre el oferente por no poseer la suficiente información para ser eficiente. Por esta razón quienes pregonan esta postura han incentivado el desarrollo y la utilización de tecnologías de información para resolver esta falla.

El gran crecimiento durante la última década en el desarrollo de las tecnologías de *Big Data* en manos del sector privado del mercado, en parte se explica por lo mencionado

anteriormente. No solo lo hacen por un incentivo de eficiencia en la generación de información con lo que luego obtienen un mayor rédito. Si no porque también se pretende solucionar las propias fallas generadas en este nuevo mercado de información. Al no surgir soluciones satisfactorias, se ha producido mayor tensión en la sociedad. De aquí que, como ha sucedido a lo largo de la historia, son las autoridades nacionales las que en muchos casos han establecido cierta regulación específica.

Ante el fracaso del mercado para poder autorregularse en particular con respecto a la privacidad de datos, en los Estados Unidos de América, el gobierno nacional ha emitido cierta regulación. El primer antecedente data de 1974, cuando se emite la "*Privacy Act*" donde principalmente se refiere a la privacidad física. No será hasta iniciado el siglo XXI, en su primera década, cuando se crea la *Federal Trade Commission*[25] (FTC) mediante la "*Act 15*". Este hecho coincide con la explosión del desarrollo y usos de tecnologías de la información en este mercado.

El accionar de la FTC es acotado. Solo en caso de incumplimiento por parte de las organizaciones en proteger a los datos de los individuos, es que podrá intervenir con peso de ley para hacer cumplir con esta obligación. Pero, aun así, la privacidad de datos personales es únicamente abordada en una relación comercial. Solo ante la existencia de una relación contractual privada, es que tiene poder de acción. Esto expone que el individuo es interpretado como un consumidor y no como sujeto de derecho humano.

Un problema que surge con esta postura de mercado, como se expuso en el apartado 1.3.2, es que no tiene en cuenta que los datos personales en ocasiones son utilizados para modificar las conductas humanas. Han sido empleados para limitar la capacidad de acción del individuo en la red o ejerciendo un control sobre el mismo. Esto es posible ya que el poder de cálculo a través de tecnología se encuentra en manos de las organizaciones y no de los individuos. Por esta razón, quien se encuentra en real desventaja como consecuencia de la asimetría de información generada, es el individuo y no la organización. A su vez, esta misma asimetría, expone el riesgo invasivo que existe sobre la privacidad del individuo como ser integrante de la sociedad (y no considerado como un consumidor).

[25] Acerca de *Federal Trade Commission*: https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement

Un caso donde puede verse representado lo anteriormente expuesto fue el escándalo por la utilización de datos personales obtenidos de la red social *Facebook*[26] para la campaña presidencial de Donal Trump en el año 2016. La agencia de *Marketing Cambridge Analytica* radicada en Londres brindaba servicios para el armado de campañas políticas bajo la promesa de "cambiar el comportamiento de la audiencia" (Kanakia, Shenoy y Shah, 2019) a través de datos. Para este fin, desarrolló un aplicativo móvil denominado "This Is Your Digital Life" a partir de la cual obtuvieron los datos de los perfiles de la red social Facebook de 270.000 usuarios. La gravedad de lo ocurrido es que lo realizaron sin el consentimiento de lo usuarios. Pero fueron un paso más allá, al recopilar los datos de 87 millones de usuarios a partir de obtener los perfiles de los amigos del grupo inicial de usuarios (Kanakia, Shenoy y Shah, 2019; ur Rehman, 2019).

A partir de los datos recopilados, *Cambridge Analytica* construye perfiles acabados de los usuarios de la red social aplicando algoritmos de *Machine Learning* para predecir cada probabilidad de voto. Esta información se utilizó para realizar una comunicación de información coercitiva (Kanakia, Shenoy y Shah, 2019). Mediante la contratación de espacio publicitario en la red social Facebook, comienza a difundir videos, fotos, noticias falsas. Esto con el fin de convencer a los usuarios de que Trump era la solución a la supuesta inestabilidad en seguridad que había en los Estado Unidos. De este modo se buscaba conquistar el voto positivo en favor del candidato presidencial.

Dado que *Cambridge Analytica* se encontraba radicada en Inglaterra, el caso comienza a ser juzgado en este país en el año 2018. Fue posible ya existía una regulación vigente en materia de la utilización de datos personales en toda la Comunidad Europea. Por otro parte, la *Federal Trade Commission* realiza una denuncia administrativa en julio 2018 acusando a la organización de haber engaño a los usuarios de la aplicación que esta había creado[27]. Como resultado, instó a la agencia de *Marketing* a que debía proteger la información personal de los usuarios o en su defecto eliminar definitivamente los datos según los términos del marco del *EU-U.S. Privacy Shield framework*. *Cambridge Analytica* no respondió a esta demanda debido a que en 2018 se declaró en quiebra y cerró sus puertas.

[26] Acerca de *Facebook* https://about.meta.com/technologies/facebook-app/
[27] Para leer el comunicado véase https://www.ftc.gov/news-events/news/press-releases/2019/12/ftc-issues-opinion-order-against-cambridge-analytica-deceiving-consumers-about-collection-facebook

Este caso permite evidenciar dos puntos que se han mencionado como riesgo asociado. Por un lado, expone como los datos pueden ser utilizados con un fin engañoso o sesgado hacia un fin en específico por las organizaciones en un entorno digital. Como fue mencionado anteriormente, se refuerza la idea de que los datos pueden ser utilizados como medios para modificar el comportamiento humano. Por el otro, que una legislación o regulación puede resultar insuficiente en la medida que solo es reactiva. Si bien existe una regulación en el caso de Europa, solo permite actuar una vez que el acto ya ocurrió. Esto no evita el daño causado.

En el caso de Estado Unidos de América, por más que existe una cierta legislación, se considera que no resulta suficiente para otorgar garantías de respeto al derecho de privacidad en base a datos personales como fue definida en este trabajo. Ello porque la responsabilidad sobre la protección de la información es puesta únicamente en manos del consumidor, una vez que el hecho ya ocurrió. Si bien puede resultar efectiva como respaldo para iniciar un reclamo, no es preventiva en el sentido de asignar responsabilidad a quienes procesan los datos.

En este sentido, se han reunido diferentes expertos y representantes de las universidades más importantes de aquel país[28] para exponer los riesgos asociados a la violación de la privacidad. Argumentando que la privacidad de datos personales es un derecho humano, enfatizan en el riesgo que conlleva la recopilación y procesamiento automático de datos. Esto con el fin de que se tomen medidas intervencionistas en este sentido. Aun así, la pregunta que surge es si este tipo de iniciativas son suficientes para alcanzar una democratización de la sociedad de la información sin la intervención de un ente regulador. Por el contrario, Zysman y Kenney (2018) exponen la necesidad de que exista una mayor regulación con el fin de evitar que se genere una mayor desigualdad en la sociedad**.** Asociado a esta perspectiva, surge la propuesta regulacionista de la Comunidad Europea.

2.2.2 El enfoque europeo

Frente a las subsecuentes consecuencias derivas de la utilización de tecnología para el procesamiento de datos personales en términos de privacidad, la Comunidad Europea decide adoptar un modelo regulacionista. Esto con el fin de intervenir en pos de brindar garantías de privacidad a los ciudadanos de los estados miembros. Para ello crea un marco regulatorio

[28] Para mayor detalle véase IMPACT 2020. Accesible a través de https://pact.mit.edu/impact-2020/

nacional, independientemente de la legislación local de cada estado que la conforma. Se trata del Reglamento General de Protección de Datos (RGPD)[29] de la Unión Europea, creado en el año 2016 y con vigencia desde 2018. Dado el avance continuo de la tecnología, ha ido sufriendo modificaciones, incorporando nuevas consideraciones.

Uno de los principios esenciales que se establece en el RGPD es que el consentimiento es la base para el tratamiento de datos personales. Esto implica que quienes recolecten y gestionen los datos personales de los individuos deberán asegurarse siempre de haber informado a los titulares. Deberán obtener su consentimiento tantas veces como sea necesario si la finalidad del uso que se le dará cambia. Esta noción se constituye en un eje fundamental para establecer un contrato social que permite generar confianza entre quienes proporcionan los datos y quienes los manejan. Pero, además, incluye otros principios sobre calidad de los datos (en términos de limitación del propósito, minimización de datos, exactitud e integridad), transparencia, acceso y rectificación, confidencialidad y seguridad (Rubinstein, 2012; Arner, Castellano y Selga, 2022).

En contraposición a la perspectiva norteamericana, la propuesta europea parte de concebir a la privacidad como un derecho humano (Arner, Castellano y Selga, 2022). Esto marca un diferencial fundamental y que se encuentra en línea con la definición propuesta de privacidad de datos personales en este trabajo. Ello porque existe un reconocimiento explícito en entornos digitales de que uno de los actores que intervienen son los ciudadanos (y no simplemente un consumidor o usuarios). Por tanto, se reconoce su derecho a la privacidad. De aquí que el individuo es puesto en el centro del procesamiento de datos. Esto permite comprender cuales son los impactos causados por el uso de tecnología para la recolección y procesamiento de sus datos.

Ahora bien, también resulta cierto, RGPD es una regulación. Entre sus lineamientos generales se establece que las organizaciones deben establecer políticas claras y de acceso disponible sobre el tratamiento que se le dará a los datos recolectados. Pero una vez más, la responsabilidad recae sobre el individuo. Si por error u omisión, el usuario no las lee, no hay garantías. Como sostiene Rubinstein (2012), diferentes estudios realizados muestran que en realidad los ciudadanos no leen de manera completa estas políticas de privacidad o en muchas ocasiones no logran comprenderlas. Esto ocasiona, que, frente a la necesidad de un

[29] Reglamento General de Protección de Datos (RGPD). Disponible en https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679

individuo de acceder a un bien o servicio, termine aceptando los términos y condiciones sin ser plenamente consciente de que es lo que realmente sucederá con sus datos.

En la dinámica de prueba y error, en posteriores modificaciones del RGPD surge la incorporación de concebir a la privacidad desde el diseño y por defecto (Agencia Española de Protección de Datos, 2019, 2020). Este enfoque significa que las organizaciones deben gestionar el riesgo asociado de manera proactiva para establecer una estrategia que incorpore a la protección de datos durante todas las etapas de su procesamiento. Con ello se busca que las responsabilidades sean asignadas por defecto, informando y formando a quienes se encuentren involucrados en cada una de estas. De esta manera, el conjunto de pautas y principios estructurados en un reglamento permite determinar qué datos son los necesarios para cada fin específico definiendo una gestión responsable. En este sentido, se reconoce que la responsabilidad del procesamiento de datos está del lado de quien lleva a cabo la tarea siendo las organizaciones.

Como se mencionó en el apartado anterior, para el caso *Cambridge Analytica*, finalmente se condenó a la agencia a pagar una suma multimillonaria por el acto cometido. Pero esta se declaró en quiebra en 2018 por lo que no cumplió con el pago. Independientemente del castigo o no que una organización pueda recibir aplicando una normativa, no previene el hecho. Al no ser de carácter preventivo sino de carácter reactivo, no evita el perjuicio – en este caso por violación de privacidad en base al uso de datos personales– generado a los individuos. Aun así, no implica que la regulación no debe existir. Es importante que exista como mecanismo de regulación del accionar de las organizaciones en el mercado. Pero no debe perderse de vista que sería importante complementar la normativa para impulsar la prevención ante el posible uso mal intencionado de los datos personales.

En este sentido es posible decir que el RGPD no establece una metodología estándar de como la gestión responsable de datos debe llevarse a cabo. Tampoco se especifican las metodologías necesarias para cumplimentar con lo en el establecido. En el caso particular de España, se crea la Agencia Española de Protección de Datos cuya misión es la de brindar asesoramiento a las diferentes organizaciones que utilizan datos, en particular personales. Pero también, a los individuos respecto de sus derechos en esta materia y la posibilidad de por ejemplo acceder a denunciar algún hecho de violación de privacidad en base a datos entre otros.

Lo que resulta interesante de dicha agencia es que ha creado diferentes guías metodológicas estandarizadas para la protección de datos personales. Incluso ha establecido un método cuantitativo robusto para aplicar[30]. Junto a este, ha establecido un diseño de estructura organizacional que permiten brindar cierta garantía de protección de datos personales (Agencia Española de Protección de Datos, 2019, 2020). Para ello cuenta con un equipo de recursos humanos especializados y multidisciplinares, que han realizado su diseño basándose en un modelo que tiene un sustento científico y académico. De este modo, se reconoce la importancia y valor de que estas existan para complementar lo que una regulación no abarca. Al mismo tiempo deja expuesto el hecho de que una regulación por sí sola no es suficiente para abarcar la problemática de privacidad en base a datos personales en el contexto de la sociedad de la información.

2.2.3 La situación en Argentina

En la República Argentina, además de lo mencionado en la Constitución de la Nación Argentina acerca del reconocimiento de la privacidad como un derecho en su artículo 19, existe la ley 25.326 de Protección de Datos Personales sancionada en el año 2000. En su primer artículo determina la necesidad de establecer:

> "…la protección integral de los datos personales asentados en archivos, registros, bancos de datos, u otros medios técnicos de tratamiento de datos, sean estos públicos, o privados destinados a dar informes, para garantizar el derecho al honor y a la intimidad de las personas, así como también el acceso a la información que sobre las mismas se registre, de conformidad a lo establecido en el artículo 43, párrafo tercero de la Constitución Nacional" (art. 1)

Particularmente, por tratamiento de datos personales refiere a:

> "… procedimientos sistemáticos, electrónicos o no, que permitan la recolección, conservación, ordenación, almacenamiento, modificación, relacionamiento, evaluación, bloqueo, destrucción…así como también su cesión a terceros a través de comunicaciones, consultas, interconexiones o transferencias." (art. 1)

[30] En particular se trata del modelo de Privacidad Diferencial que ya ha sido probada su efectiva en el año 2008 por la Dra. Cinthya Dwork y que será presentado en el capítulo 3 del presente trabajo.

Es importante remarcar que la ley 25.326 en ningún momento menciona el término privacidad de datos personales. Esto marca la desactualización que posee ante el avance de las tecnologías de la información y los cambios sociales que ello ha producido durante las dos décadas posteriores a su promulgación. Si bien han surgido diferentes solicitudes de parte de los representantes de los ciudadanos para darle tratamiento en el congreso nacional sobre posibles reformas[31], aún no se ha logrado que esto se lleve a cabo.

Además, la Ley establece que se enmarca únicamente dentro del artículo 43 de la Constitución Nacional. En lo que refiere a datos particularmente, dicho artículo establece que:

> "Toda persona podrá interponer esta acción para tomar conocimiento de los datos a ella referidos y de su finalidad, que consten en registros o bancos de datos públicos, o los privados destinados a proveer informes, y en caso de falsedad o discriminación, para exigir la supresión, rectificación, confidencialidad o actualización de aquéllos. No podrá afectarse el secreto de las fuentes de información periodística." (Art.43)

De lo aquí expuesto es posible interpretar que la responsabilidad del control de los datos recae sobre el individuo. Si bien se explicita sobre el derecho a reclamar, no provee un enfoque de prevención al no asignar responsabilidad de brindar garantía a quienes utilizan los datos. En cierto modo, se presenta una situación similar como fue mencionado para el caso del RGPD en el apartado anterior.

A pesar de la existencia de esta legislación ocurrieron diferentes hechos de violación a la privacidad en base a la utilización de datos, como se mencionó en la introducción del presente capitulo. Por ejemplo, se encuentra el caso de la suspensión del sistema de vigilancia por reconocimiento fácil de prófugos del Gobierno de la Ciudad Autónoma de Buenos Aires. La razón por la que la justicia emitió esta orden fue por constatar que en el Ministerio de Justicia y Seguridad de la Ciudad Autónoma de Buenos Aires se habían utilizado datos personales (biométricos) obtenidos de RENAPER (Registro Nacional de las Personas) de individuos que no eran prófugos. De esta manera, se evidenció una violación en el acceso a información personal. Mientras en la base de prófugos hay un registro de 40

[31] En línea: https://www.argentina.gob.ar/aaip/datospersonales/proyecto-ley-datos-personales

mil individuos, las consultas a la base de datos de RENAPER para la obtención de los datos biométricos por parte de dicho Ministerio superaron los 16 millones (Ámbito.com., 2023).

Este último hecho, también es un claro ejemplo de cómo en una organización se utilizan datos propios en combinación con otros provenientes de fuentes externas para la identificación de individuos incurriendo en la violación de la privacidad. Se incurre en un incumplimiento de la ley porque se ha aprovechado la posibilidad de acceso a los datos personales para el cruce de la información con fines desconocidos y por fuera de un requerimiento judicial. Si existiera una orden judicial, es un caso de excepción que si está previsto por la ley anteriormente mencionada. Pero, además, expone como la legislación no previene el uso mal intencionado de los datos para evitar el perjuicio a los individuos, sino que se aplica una vez que el hecho ya ocurrió. A su vez, pone de relieve la necesidad de contar con un enfoque de prevención de forma tal que se definan métodos para llevar a cabo una gestión responsable de la información en las organizaciones públicas.

En cambio, en el caso de acceso no autorizado a la base de datos de PAMI (Instituto Nacional de Servicios Sociales para Jubilados y Pensionados) para el robo de datos personales de miles de jubilados (elcronista.com, 2023), pone de relieve otro aspecto. En este caso asociado a la responsabilidad en la gestión de los datos en un contexto organizacional. En esta ocasión, la falta de protocolos informáticos de seguridad para evitar que cualquier usuario acceda a la base de datos personales también resulta necesario y parte integrante de un diseño responsable de gestión de datos. De aquí que, los frentes sobre los cuales aplicar controles estrictos en pos de construir garantías para una protección de datos personales pueden resultar muy diversos. Pero, principalmente, la idea de prevención dentro de la organización antes que un accionar reactivo toma más fuerza. Para la creación de conciencia sobre protección de datos personales requerirá de la asunción de responsabilidad y formación de quienes se encuentren involucrados.

Por otra parte, en Argentina existe la Subsecretaría de Tecnología de la Información, que busca promover un uso responsable de la tecnología aplicada a datos principalmente en el sector público. Pero esta no cuenta con una metodología estandarizada sobre el tratamiento de datos personales, como es el caso de la Agencia Española mencionada en el apartado anterior. No obstante, en junio de 2023, ha emitido la Disposición 2/2023 sobre un marco

general de recomendaciones para una Inteligencia Artificial Fiable[32] en línea con los principios establecidos por la UNESCO[33] y la OCDE[34] al respecto. A partir de esta se busca promover la protección de datos personales a la hora de implementar modelos de inteligencia artificial desde una perspectiva ética y de derechos humanos. Si bien se considera que es un buen primer paso para avanzar en la asignación de responsabilidad en el tratamiento de la protección de datos personales aún queda mucho camino por recorrer.

Otra organización pública del estado que tiene vinculación con promover la protección de datos personales es la Agencia de Acceso a la Información Pública (AAIP)[35]. Esta dispone de un registro de infractores de la Ley 25.326 desde el año 2016. Obteniendo los datos contenidos en el mismo con *Python* (ver código en Apéndice – sección "Apéndice A0 Registro de infractores Ley 25.326"), se puedo determinar que, a octubre de 2024, 92 organizaciones han violado dicha Ley al menos una vez. Es de destacar que entre dicha información en ningún caso se encuentran organizaciones públicas del estado, ni nacionales, ni provinciales, ni municipales. Solo se trata de organizaciones de carácter privado, entre las que se encuentran grandes actores del mercado como las compañías de telefonía.

Frente a este escenario, es posible decir que en Argentina si bien existe una legislación vigente sobre la privacidad de datos, no resulta suficiente por si sola más allá de su desactualización. Los hechos ocurridos como los anteriormente mencionados ponen en evidencia la necesidad de gestionar responsablemente los datos personales en un contexto organizacional para brindar garantía de privacidad a sus titulares. Además, porque tampoco se cuenta con un estándar metodológico para dar tratamiento a la protección de datos personales que asigne responsabilidad sobre quienes lleven adelante su procesamiento. Su existencia se considera relevante para contribuir un sistema democráticamente más justo en materia de privacidad de datos personales en los términos que fue definido en este trabajo.

2.2.4 Comparativa y resumen

En base a los expuesto en los apartados previos, en este apartado realiza un cuadro comparativo a modo de resumen para exponer los resultados obtenidos sobre el enfoque y tratamiento de la privacidad de datos personales en los tres países analizados. De este modo,

[32] En línea: https://www.boletinoficial.gob.ar/detalleAviso/primera/287679/20230602
[33] En línea: https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
[34] En línea: https://oecd.ai/en/ai-principles
[35] En línea: https://www.argentina.gob.ar/aaip

se busca visualizar sintética y específicamente las perspectivas planteadas y la necesidad en el caso argentino de poder contar con al menos una metodología estandarizada en materia de protección de datos personales.

Cuadro 1: Resumen de enfoques de la privacidad de datos personales (PDP)

País	Enfoque PDP	Individuo	Posee algún tipo de legislación vigente PDP	Es suficiente	Posee Metodología estandariza PDP
Estados Unidos de América	Mercado	Consumidor	Si	No	No
Comunidad Europea	Derecho humano	Ciudadano	Si	No	Si
Argentina	Derecho humano	Ciudadano	Si	No	**No**

Fuente: elaboración propia

En el cuadro 1 se puede observar que el enfoque norteamericano solo asigna la responsabilidad en el individuo al ser considerado un consumidor. Se considera que no resulta adecuado para abordar el tratamiento de la privacidad de datos personales de manera preventiva, como se ha expuesto en los apartados anteriores. En cuanto a la propuesta europea, resulta un buen ejemplo de regulación avanzada pero que aún no logra ser suficiente en la prevención dado su carácter reactivo. En el caso argentino, si bien se parte de la concepción ética y de derecho humano sobre la privacidad, la ley vigente no se encuentra adaptada al contexto actual de evolución tecnológica. Pero, además, solo es de carácter reactivo.

Particularmente, en el caso de Argentina, por un lado, surge la necesidad de una actualización de la ley 25.326 de Protección de Datos Personales debido a los avances tecnológicos para la obtención de información en base a datos. Por el otro, resultaría importante que también se revise la asignación de responsabilidades en materia de regulación. Aun así, posiblemente no resulte suficiente. Dado el poder contraído por las organizaciones públicas sustentado en una asimetría de información, el ciudadano queda en una situación de vulnerabilidad respecto de su privacidad. Por esta razón, resultaría acorde que se elabore cierto marco basado en asignación de responsabilidad por parte de la Subsecretaría de Tecnología de la Información.

Para que en la sociedad argentina se pueda alcanzarse una cultura de responsabilidad en un contexto organizacional para el tratamiento de datos personales, resultaría adecuado que desde las esferas públicas se promueva el desarrollo de metodologías. Para ello resultará

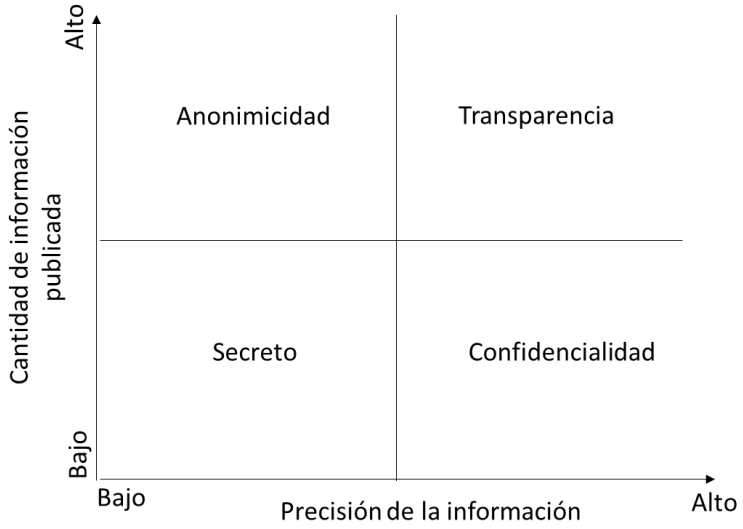
adecuado evaluar sus fundamentos científicos y realizar pruebas para la validación de resultados. De esta manera, se podrá acompañar a las diversas organizaciones en la adopción de esta propuesta, así como también en la adaptación a su contexto en específico y no solo tratarse de una imposición regulatoria.

En este sentido, incorporar especifica y explícitamente la concepción de la privacidad desde el diseño y por defecto en materia de privacidad de datos personales, puede resultar acorde. Esto porque permitirá diseñar una gestión de los sistemas de datos en contextos organizacionales desde una perspectiva de responsabilidad. La aplicación de métodos cuantitativos y asignación de responsabilidades resultarán cuanto menos esenciales. A continuación, se especifica sobre este concepto para alcanzar su comprensión.

2.3 Privacidad desde el diseño y por defecto

La noción de privacidad desde el diseño y por defecto, inicia originalmente con la ética de la información desde la perspectiva de las tecnologías de la información y comunicación (TIC). Con ello se incorpora a la ética en una disciplina técnica. Para poder lograrlo, esta perspectiva parte de cuatro términos claves que fueron definidos en el aparatado 2.2.1, siendo el anonimato, el secreto, la transparencia y la confidencialidad. A partir de relacionarlos con respecto a la cantidad de información personal recolectada, surge el riesgo de identificación de individuos en base a datos personales.

Figura 1: Riesgo de identificación individual en base a datos



Fuente: tomado y traducido de Smith, Dinev, & Xu (2011)[36]

[36] Quienes a su vez extraen el esquema de Zwick y Dholakia (2004)

En la figura 1 es posible observar cómo se produce la identificación o reconocimiento de un individuo digitalmente. Viene dada por la cantidad y exactitud de la información personal recopilada. A mayor cantidad de información a ser publicada (o utilizada), será necesario lograr una mayor anonimicidad para evitar la identificación individual. En consecuencia, si se quiere recopilar más información será necesario lograr una mayor confidencialidad y transparencia de procesos para alcanzar mayor precisión. Con ello se logra exponer la dificultad que se presenta en el uso de datos personales en términos de riesgo asociado a la identificación de un individuo derivado del uso de la tecnología aplicadas a datos.

A medida que las organizaciones intensifican el uso de las tecnologías del *Big Data,* se comienza a obtener cada vez mayor capacidad de recolección de datos y la posibilidad de obtener una mayor exactitud en la identificación de individuos. En consecuencia, el riesgo de identificación de un individuo comienza a ser aún mayor. En este contexto, desde la perspectiva de las TIC, se comienza a hablar del valor de la sensibilidad del diseño (*Value Sensitive Design*) de los sistemas de datos (Friedman, Kahn, Borning & Huldtgren, 2013). Con ella se refiere a la necesidad de diseñar sistemas tecnológicos de información pero que tengan en cuenta los valores humanos y principios éticos de una manera integral.

Desde esta perspectiva, se reconoce la relevancia de la sensibilidad de los datos recopilados y procesados ya que puede afectar la privacidad de los individuos. El uso de identificadores directos o indirectos definidos en el aparatado 2.1.1 permiten construir perfilamientos acabos de los individuos (Custers, 2013; Haskel y Westlake, 2017; Ghandy Jr., 2021). A partir de ello, el riesgo generado en términos de quebrar la barrera de acceso a las cuestiones que el individuo decide mantener como privadas, comienza a ser muy alto. Esto porque el desarrollo tecnológico brinda cada vez mayor potencialidad de procesamiento en las organizaciones. De aquí que, desde la ética y los valores humanos, esto comienzan a ser cuestionado.

A partir de ello, y con el transcurso de los años, la gestión responsable de los procesos de datos organizacionales comienza a ser demandada por la sociedad. Junto con ciertos principios éticos establecidos por Cavoukian (2011), dan lugar a la idea de privacidad desde el diseño y por defecto para ir posicionándose como una gestión responsable de privacidad. Esto porque logra incorporar como un componente relevante a la responsabilidad sobre los procesos de datos por parte de las organizaciones. Ya no que sea solo una responsabilidad de los individuos como propone la mayoría de la legislación.

La privacidad desde el diseño y por defecto significa entonces gestionar el riesgo asociado a la privacidad de datos personales de manera proactiva para establecer una estrategia que incorpore la protección de estos datos durante todas las etapas de su procesamiento en las organizaciones (AEPD, 2019, 2020). Por un lado, la proactividad requerirá la implementación de metodologías en el diseño de los sistemas de datos para reducir el riesgo. Por el otro, las responsabilidades serán asignadas por defecto, informando y formando a quienes se encuentren involucrados en cada una de estas. A continuación, se desarrolla sobre cada uno de estos aspectos.

2.3.1 Privacidad de datos personales por defecto

A partir de incorporar la ética a los sistemas de datos implica que las organizaciones creadoras y poseedoras de la tecnología construyan una responsabilidad en la gestión del riesgo asociado a la identificación de individuos en base a datos. Esto porque resulta en un factor clave para la construcción de una gestión responsable de la privacidad de datos personales como se planteó en la definición realizada en el apartado 2.1.1. En este sentido, se requiere considerar cuatro aspectos principales: la cantidad de datos personales, su tratamiento, el modo de conservación y la accesibilidad (AEPD, 2019, 2020).

La cantidad de datos personales, como se puedo observar en la figura 1 del apartado anterior, que son recolectados implica tanto aspectos cuantitativos como cualitativos. Quien lleve adelante esta tarea dentro de la organización, deberá tener en cuenta el volumen de datos tratados para discernir entre los realmente necesarios y los que no (AEPD, 2019, 2020). Pero también deberá considerar el tipo de dato requerido. No todos los datos personales necesariamente deben ser recopilados, utilizados o expuestos o incluso podrán ser anonimizados de forma tal de protegerlos.

El tratamiento de datos personales implica que su uso deba limitarse a los estrictamente necesario (AEPD, 2019, 2020). Es decir, dada las diferentes etapas por las que atraviesan, deberá tenerse en cuenta que en cada una de ellas se realicen las operaciones únicamente necesarias para el cumplimiento de cada una. De este modo se podrá evitar su reutilización ya que puede derivar en un uso diferente por el cual fueron originalmente recolectados (Steinmann, Matei y Collmann, 2016; Gandy, 2021). Pero también, será necesario administrar el tiempo de conservación de estos.

La conservación de datos personales requerirá de una atención particular. Su perdurabilidad debe ser limitada (AEPD, 2019, 2020) para lograr el principio de minimizar riesgos de uso. En ocasiones, durante una fase de análisis, se desarrollan procesos para la generación de nuevos datos. Cuando esto sucede, los datos son derivados o generados de una recombinación. Entonces, se deberá tener en cuenta no conservarlos indefinidamente. En su lugar es recomendable resguardar los procesos generadores para que puedan ser repetidos en caso de ser necesario. En el caso particular de Argentina, además, la Ley 25.326 limita el tiempo de conservación a un cierto período, por lo cual también la responsabilidad es en términos de cumplir con la legislación vigente.

La responsabilidad también involucra la accesibilidad a datos personales en cada una de las fases del ciclo de vida en la organización (AEPD, 2019, 2020). Desde una perspectiva técnica, será necesario establecer criterios de perfiles minuciosos para el acceso a los datos limitado a la necesidad específica. Por un lado, esto implica que no cualquier persona pueda acceder a todos los datos, sino que solo a los necesarios. Por el otro, resultará adecuado establecer protocolos de control para publicación o transferencia de datos personales. La responsabilidad en este sentido es evitar fuga de información o un uso mal intencionado, además de cumplir con lo establecido por Ley como es el caso argentino.

De esta manera, la responsabilidad por defecto será transversal a toda la organización que entre otras cosas requerirá informar y formar entorno a la privacidad a todos aquellos que se encuentren involucrados en los sistemas de datos personales. A su vez, esto implica la generación de una cultura del compromiso ético, impulsada desde la organización para la construcción de la protección de datos personales. Pero esto no será posible de alcanzar si al mismo tiempo no se lleva adelante la incorporación de la privacidad desde el diseño de los sistemas de información.

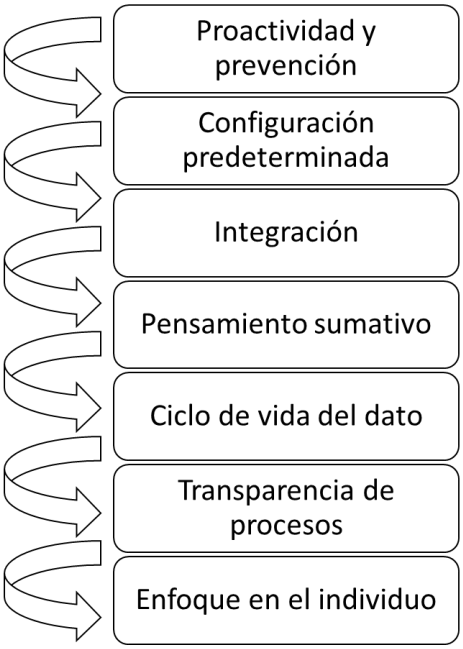
2.3.2 Privacidad de datos personales desde el diseño

Para lograr una gestión responsable de datos personales, resulta necesario mitigar el riesgo asociado a la privacidad. Para ello será necesario considerarla desde la fase del diseño de los sistemas de datos en las organizaciones. Desde una perspectiva sociológica como fue mencionado al inicio de este capítulo, implica posicionar al ciudadano en el centro del procesamiento de datos para evaluar el impacto causado en términos de su derecho a la privacidad. Pero ello solo no resultará suficiente. También, como se mencionó en el apartado

2.3, es necesario considerar la sensibilidad del valor de los datos desde la perspectiva ética de las tecnologías de la información.

En este sentido, Cavoukian (2011) ha sintetizado la perspectiva ética de las tecnologías de la información en siete principios fundacionales de la privacidad para su incorporación en los sistemas de datos. Estos se han convertido en una referencia mundial a tal punto que han sido incorporados en la regulación canadiense pero también en la regulación europea. Estos principios pueden observarse en la Figura 2 a continuación.

Figura 2: Principios fundacionales para el diseño de la privacidad



Fuente: elaboración propia en base a Cavoukian (2011)

En la figura 2, la proactividad y prevención en materia de privacidad de datos personales, implica tener una concepción proactiva y preventiva antes que reactiva y correctiva (Cavoukian, 2011; AEPD, 2019, 2020). Ante la sensibilidad de la información resulta imperioso que sea gestionada adecuadamente en los procesos de datos con el fin de evitar o minimizar los riesgos asociados a violación de la privacidad. Esto podrá ser alcanzado mediante la implementación temprana de metodologías cuantitativas. A partir de ello, la incorporación de la privacidad en el diseño de los sistemas permitirá que quede establecida e integrada como una configuración predeterminada. De este modo, los datos personales serán protegidos de forma automatizada a través de implementar los métodos necesarios a través de la propia tecnología. Pero esto solo no será suficiente. También será necesario que

se establezca limitación en el uso de los datos, así como una adecuada conservación, y restricciones de acceso acorde a la necesidad de cada usuario en la organización.

Para evitar caer en una discusión de privacidad versus oportunidad de uso en la organización, resulta fundamental construir un pensamiento sumativo en un sentido éticamente responsable (Cavoukian, 2011; AEPD, 2019, 2020). Esto facilitará que la privacidad entendida como un derecho humano sea comprendida a partir de la sensibilidad de los datos personales. Teniendo en cuenta que los datos atraviesan diferentes etapas durante su ciclo de vida en la organización como fue mencionado en el apartado 1.2.2, será necesario establecer garantías de privacidad dentro de todo el ciclo. No solo alcanza con definirla únicamente en el diseño previo a la implementación, sino que también deberá garantizarse durante todas las etapas posteriores. Ello requerirá de un análisis de todas las etapas que atraviesa el dato, implementando en cada una las medidas adecuadas.

La transparencia resulta un principio fundamental ya que con este se busca mostrar la responsabilidad en el tratamiento de los datos personales de forma tal de generar confianza con el individuo (Cavoukian, 2011; AEPD, 2019, 2020). Para lograrlo será necesario hacer públicas las políticas de privacidad que establece la organización el marco de los principios antes mencionados como así también en base a la regulación vigente. Para lograrlo será necesario conocer el nivel de comprensión de los individuos, es decir, si fuese necesario también construir conciencia de privacidad en ellos además de que sean lo suficientemente claras. La comunicación transparente y de fácil entendimiento debe ser la clave en este proceso, para lograr centrar la atención en el individuo.

Ahora bien, los principios mencionados son enunciaciones que sirven como guía en una organización para alcanzar una gestión éticamente responsable de privacidad de datos personales. Por esta razón resultan cuanto menos importantes considerarlos, pero no suficientes. Para alcanzar una eficiencia y efectividad en la protección de datos personales, resultará necesario seleccionar, evaluar e implementar una metodología que brinde garantía de privacidad.

Un método frecuente como primera aproximación a garantizar la privacidad, es la anonimización de datos (AEPD, 2019, 2020). Esta consiste en la generación de códigos alfanuméricos sobre un valor de dato dado. Pero resulta limitada ya que no es posible anonimizar todos los identificadores debido a que inhabilita su posterior uso para construir

información de análisis. Si bien puede utilizarse para ocultar identificadores directos cuando se decide publicar datos como el CUIT, los identificadores indirectos en muchas ocasiones aportan valor analítico. En este sentido, será necesario utilizar otra metodología que no ocasione este inconveniente.

Una metodología que se ha demostrado que brinda garantía de privacidad, permitiendo medir su efectividad, es la Privacidad Diferencial. Se trata de una metodología cuantitativa que consiste aplicar ruido aleatorio a los datos personales con el fin de preservar su privacidad sin que se inhabilite la utilidad de estos (Dwork, 2008; Dwork y Roth, 2014). Este método será abordado en el capítulo 3 a continuación para lograr su comprensión y luego continuar con una demostración de su aplicación.

Conclusión del capítulo

En el presente capítulo, se ha elaborado una definición de la privacidad de datos personales. Ha sido definida como la posibilidad de que un individuo generador de datos al interactuar con una organización en un contexto de red digital pueda obtener garantías suficientes de que sus datos están siendo debidamente protegidos. Por un lado, esta definición reconoce el derecho humano a la privacidad que poseen los individuos permitiendo que puedan decidir si aceptan o no que una organización procese sus datos personales. Por el otro, implica que las organizaciones públicas y privadas deben asumir una gestión responsable que garantice la protección de los datos personales que recopilan y utilizan con el fin de brindar garantías de privacidad a los ciudadanos.

La privacidad como derecho humano encuentra su origen en 1948 en la Declaración Universal de Derechos Humanos realizada por la Organización de Naciones Unidas. Pero esta perspectiva concibe a la privacidad en un sentido físico. Gran parte de la legislación vigente en los países se sustenta en esta concepción. Pero no será hasta finales del siglo XX y comienzo del siglo XXI, cuando se comienza a pensar a la privacidad en términos de perjuicio por invasión a la intimidad como consecuencia de la obtención de conocimiento en base al procesamiento de datos personales. Facilitado por el avance de los desarrollos de las tecnologías del *Big Data*, comienza a cuestionarse el impacto causado en torno al derecho a la privacidad.

A partir de ese entonces, y como fue mencionado en el capítulo 1 del presente trabajo, la sociedad se configura como un híbrido colectivo entre organizaciones, individuos y

tecnología en red. Ante el mayor poder de las organizaciones, el ciudadano queda expuesto a una situación de vulnerabilidad en términos de privacidad. Esto debido a la existencia de un procesamiento de datos omnipresente y continúo creado y controlado por las organizaciones. Es entonces que el impulso del desarrollo tecnológico en manos del mercado comienza a ser cuestionado. La concepción de los individuos como un consumidor expone la vulnerabilidad de los derechos de los ciudadanos. Particularmente, la preocupación acerca de las consecuencias sobre la privacidad en base a datos personales comienza a incrementarse.

Como primera respuesta, los países crean regulación específica sobre la problemática de privacidad de datos personales, pero con diferente enfoque según el sistema de configuración social y cultural. A partir de ello surgen dos posturas bien diferenciadas. Por un lado, la perspectiva norteamericana que basa su concepción en el paradigma del libre mercado. A partir de ello concibe al individuo como un consumidor brindándole cierta posibilidad de reclamo. La legislación es utilizada como medio para poder reclamar en caso de que su intimidad a partir del uso de datos personales por parte de alguna organización le cause cierto perjuicio. Esta perspectiva no resulta preventiva ni proactiva, sino que carga la responsabilidad de accionar en el individuo una vez que el incidente ya ocurrió.

Por el otro, se encuentra la regulación de la Comunidad Europea. A partir de la creación del Reglamento General de Protección de Datos en 2016, se incorpora a la privacidad de datos como un derecho humano. Esto marca un diferencial fundamental con respecto a la perspectiva norteamericana y que se encuentra en línea con la definición propuesta de privacidad de datos personales realizada en este trabajo. Ello porque existe un reconocimiento explícito en entornos digitales de que uno de los actores intervinientes son los ciudadanos y no simplemente un consumidor o usuario y, por lo tanto, posee derecho a la privacidad. En este sentido, el individuo es puesto en el centro del procesamiento omnipresente de datos en red reconociendo el impacto que le causa en términos de lo que decide mantener como privado. Pero aun así cierta responsabilidad sigue recayendo en el individuo ya que recién frente a un reclamo por un hecho ocurrido se actúa en consecuencia.

Argentina, cuenta con la Ley 25.326 de Protección de Datos Personales creada en el año 2000 y sin actualización. Si bien otorga cierta garantía legal al ciudadano, resulta insuficiente porque en gran parte de su contenido se lo concibe como usuario o consumidor al estilo norteamericano. Aun así, incorpora ciertos conceptos de la visión europea reconociendo a la

privacidad como la privacidad de datos personales desde una perspectiva ética del derecho humano en línea con la definición realizada en este trabajo. No obstante, un diferencial marcado con la perspectiva europea es que en Argentina no se cuenta con una metodología estandariza de protección de datos personales en organizaciones como en el caso de por ejemplo España. La propuesta española de concebir a la privacidad desde el diseño y por defecto en los procesos de información en contextos organizacionales, se ha convertido en una referencia mundial.

Frente a la insuficiencia de la regulación existente en el caso argentino, se considera necesario incorporar también la perspectiva ética de la tecnología que permite centrar la atención en la sensibilidad de los datos personales. Esto dará lugar a la concepción de la privacidad desde el diseño y por defecto, de forma tal de poder construir una propuesta metodológica estandarizada que la abarque en el marco de la definición establecida en este trabajo. Esta propuesta significará gestionar el riesgo asociado a la privacidad de datos personales de una manera proactiva y preventiva antes que reactiva. La importancia de ello es que busca reconocer la responsabilidad que deben asumir las organizaciones a la hora de diseñar sus procesos de datos en términos de brindar garantías.

En el marco de la propuesta de la privacidad desde el diseño y por defecto, la protección de datos personales deberá ser considerada en cada una de las etapas del procesamiento. Junto a la consideración de ciertos principios éticos definidos en el último apartado del presente capítulo, se podrá reconocer y asumir la responsabilidad necesaria por parte de las organizaciones. A su vez, también resultará importante que la autoridad responsable establezca una metodología estandariza para realizar una evaluación cuantitativa de la protección de datos personales. Esto contribuirá a que las organizaciones tengan un marco de acción concreto para establecer una medida de control cuando se utilizan este tipo de datos.

En este trabajo se propone utilizar la metodología de la Privacidad Diferencial. Dado que se ha demostrado que brinda cierta garantía de privacidad, permitirá medir su efectividad para la protección de datos personales. Se trata, de una metodología cuantitativa que consiste aplicar ruido aleatorio (distorsión) a los datos personales con el fin de preservar su privacidad sin que se inhabilite la utilidad de estos y que permite medir su eficacia. Esta será presentada y desarrollada en el capítulo 3 a continuación mediante un caso de aplicación.

Capítulo 3: La metodología de la Privacidad Diferencial

Introducción

En base a lo expuesto en los capítulos anteriores de presente trabajo, surge un riesgo asociado a la privacidad en torno a la utilización de datos. Dada la recopilación y procesamiento de datos personales llevado a cabo en las organizaciones, surge la posibilidad de identificación individual lo que corrompe el derecho a la privacidad. Pero también que la información generada puede ser utilizada para la modificación de acciones posibles colocando al individuo en una situación de desventaja. Frente a ello, y ante el carácter reactivo de la regulación, se argumentó sobre la necesidad de que las organizaciones adopten una actitud preventiva dentro de un marco de responsabilidad.

La acción preventiva se asocia a la mitigación de riesgos asociados al uso de datos personales y puede ser alcanzada a través de utilizar un método cuantitativo. Por un lado, su aplicación deberá brindar la posibilidad de obtener cierta garantía de protección de datos y evitar la identificación de individuos. Por el otro, que siga permitiendo la utilización de los datos para la generación de información. Para ello sería necesario poder medir su efectividad. De este modo, resulta importante que permita encontrar un equilibrio entre la necesidad de uso y la protección. Así podrá contribuirse responsablemente a que la interacción entre individuos y las organizaciones en un espacio digitalizado se lleve a cabo dentro de un marco de responsabilidad.

En la búsqueda del equilibrio entre la protección de datos personales y la necesidad de uso de los datos, la responsabilidad también involucra que sea haga de una manera beneficiosa para los individuos. Cuando se trata de organizaciones estatales, resulta necesario utilizar esta información para la construcción de políticas públicas que contribuyan al bienestar de la sociedad. En este sentido, la propuesta de la metodología de la Privacidad Diferencial podría resultar adecuada.

La Privacidad Diferencial es una metodología cuantitativa con fundamento probabilístico matemático que permite proteger datos personales en una base de datos sin invalidar su utilidad (Dwork, 2008). A su vez permite decidir el nivel de protección que se desea aplicar o que resulta necesario para brindar garantías de protección. Metodológicamente, se trata de aplicar ruido aleatorio a los datos personales al momento de obtenerlos para luego implementar técnicas para su procesamiento. Como resultado se obtendrá que su distribución

no cambiará a nivel agregado, es decir, en el total de la población utilizada. Pero requerirá de la evaluación de un costo, que vendrá dado por cuanta información real se está dispuesto a no publicar o utilizar a cambio de obtener protección.

En este contexto, surge el siguiente interrogante: ¿qué nivel de efectividad tiene la aplicación de privacidad diferencial para la protección de datos personal? Para responder esta pregunta el objetivo a desarrollar en el presente capítulo consiste en presentar el modelo de la privacidad diferencial y sus fundamentos para luego realizar una aplicación de caso que aproxime a una respuesta local. Se comenzará por presentar el concepto de Privacidad Diferencial. Dado su carácter técnico, también se presentarán todos los conceptos técnicos involucrados para lograr una comprensión acabada. Luego se presentará el modelo matemático probabilístico que la fundamenta. Posteriormente, se construirá una base de datos personales y se realizará una identificación de individuos. Finalmente, se aplicará el modelo de privacidad diferencial presentado y se evaluará su efectividad en términos de evitar la reidentificación de individuos como método para obtener protección de datos personales. De este modo podrá evaluarse si es viable de ser considerado para la construcción de una gestión responsable que brinde garantías de privacidad frente a ciertos atributos personales como los contenidos en la base de datos construida.

A fin de explicitar la metodología de la privacidad diferencial, el presente capítulo se divide en cuatro secciones. En la primera sección se desarrolla la conceptualización de la privacidad diferencial desde una perspectiva teórica. Para ello, se toma el concepto de privacidad diferencial que fue propuesto y desarrollado por la autora Cynthia Dwork en 2008. A partir de esta definición, se amplía la exposición para mostrar la lógica involucrada en el método. Finalmente se concluye acerca de su definición en relación con la protección de datos personales.

En la segunda sección, se presentan tres conceptos involucrados en la modelización de la privacidad diferencial que luego permitirán realizar la exposición y comprensión del modelo. Para ello se comienza por abordar el concepto de función o algoritmo aleatorio. Luego se continúa por el análisis de comportamiento de la distribución de probabilidad de Laplace. Finalmente se presenta el mecanismo de Laplace que es parte constitutiva del modelo propuesto por la autora anteriormente mencionada para la generación de ruido aleatorio.

En la tercera sección, se aborda la presentación del modelo de privacidad diferencial. Para ello en primer lugar, se expone la modelización matemática creada por su autora Cynthia Dwork. Luego, se continúa con la interpretación necesaria teniendo en consideración los conceptos previos presentados. Por último, se centra la explicación sobre el parámetro del modelo que permite evaluar el nivel de protección sobre datos. Este será la clave para la definición de una estrategia responsable de protección de datos como método a aplicarse en un contexto organizacional.

Finalmente, en la cuarta y última sección se lleva a cabo una aplicación práctica del modelo sobre los atributos personales contenidos en la base de datos construida. En el primer subapartado se explicita sobre la construcción de la base de datos personales con la cual se trabajará. A continuación, se realiza una descripción de los datos finalmente obtenidos y se explicita cómo es posible identificar de manera directa individuos. Finalmente, se muestra cómo es posible llevar a cabo una reidentificación indirecta en base a la combinación de atributos personales.

En el segundo subapartado se desarrolla un algoritmo para aplicar la metodología de la Privacidad Diferencial. En primer lugar, se detalla la estructura del algoritmo programado con *Python*. Luego, para ciertos parámetros, se lo aplica sobre un atributo personal contenido en la base de datos previamente construida. Finalmente, en base a los resultados obtenidos, se evalúa la efectividad de la metodología en términos de si impide la identificación de un individuo en la base de datos.

Por último, en el tercer subapartado, en base los resultados obtenidos previamente se lleva a cabo la aplicación de ruido aleatorio por el mecanismo de Laplace con *Python*. Para ello, en primer lugar, se detalla la estructura del algoritmo desarrollado que permite llevar a cabo su aplicación. A continuación, se aplica en el entorno *Google Colaboratory* sobre toda la base de datos. Finalmente, en base a los resultados obtenidos, se evalúa si el método resulta efectivo para la protección de los datos personales contenidos en la base utilizada.

3.1 El concepto de Privacidad Diferencial

La privacidad diferencial es una metodología cuantitativa para proteger los datos personales que se encuentran en una base de datos. Fue introducida y desarrollada por Cynthia Dwork en 2008 en su trabajo "*Differential Privacy: A Survey of Results*". Para evitar la identificación de individuos, el objetivo propuesto por la autora es implementar un método

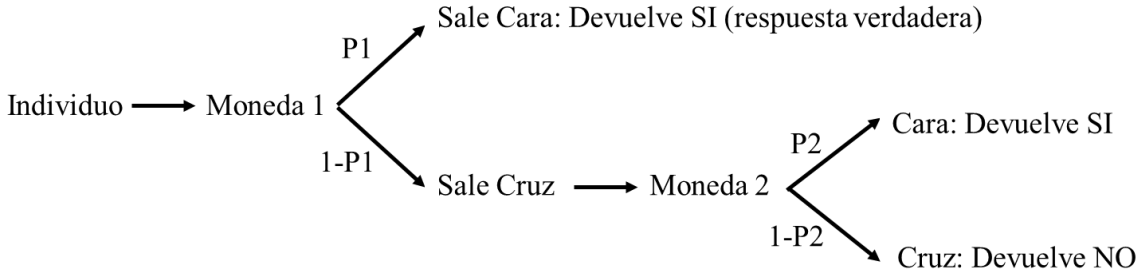
que distorsione los identificadores directos e indirectos contenidos en una base de datos. Con ello busca obtener información sobre el comportamiento de una población, pero no así de un individuo en particular evitando violar su privacidad. A su vez implica que, si los datos son recombinados con otros provenientes de fuentes externas a la organización, no podrá realizarse una reidentificación de individuos.

Siguiendo a la autora Cynthia Dwork (2008), metodológicamente la privacidad diferencial consiste en aplicar ruido aleatorio sobre los datos personales al momento de implementar técnicas para su procesamiento. Como resultado se obtendrá una cierta distorsión en los mismos sin afectar su utilidad posterior. Esto vendrá dado por garantizar que la distribución de las variables involucradas no cambiará en el total de la población. De este modo, el método brinda la posibilidad de contar con un método matemático probabilístico riguroso para hacer frente a la posibilidad de re identificar al individuo titular de los datos. Pero requerirá de la evaluación de un costo. Esta consiste en determinar cuanta información real se está dispuesto a no publicar o utilizar a cambio de obtener protección, es decir, privacidad.

Para poder explicar la lógica de funcionamiento del método de la privacidad diferencial, se ejemplifica a través de la respuesta aleatoria. Esta última fue desarrollada por las ciencias sociales (Dwork, 2008) para recopilar información estadística sobre alguna temática tabú. Por ejemplo, supongamos que se quiere analizar los resultados de una encuesta sobre una temática sensible. Las respuestas son del tipo sí o no, considerando que los encuestados (titulares de los datos) no permiten el acceso directo a la información real brindada.

Frente a este escenario es posible aplicar privacidad diferencial para evitar exponer cierta información que el individuo encuestado decide mantener como privada (Dwork, 2008; Ghandy Jr., 2021). Suponga que el proceso contiene la siguiente lógica. Se lanza una moneda que representa la aleatoriedad. Si sale cara devuelve la respuesta verdadera SI. Si no, se lanza una segunda moneda, y si sale cara devuelve como respuesta SI mientras que si sale cruz devuelve como respuesta NO. A continuación, se puede observar este proceso de manera esquemática.

Imagen 0: Esquema de obtención de respuesta aleatoria



Fuente: elaboración propia en base a Dwork (2008)

En el esquema contenido en la imagen 0, los datos de cada individuo estarán protegidos con una negación plausible mediante el lanzamiento de una moneda (aleatoriedad) (Dwork, 2008). De este modo, si se quiere conocer la probabilidad obtener la respuesta SI una vez aplicada la privacidad diferencial (PD) sobre los datos, se deberán considerar las siguientes probabilidades condicionales:

$$P(SI/está\ protegido\ por\ PD) = P1 + [(1 - P1) * P2] \quad (1)$$

siendo (1) la probabilidad de obtener como respuesta SI dado que los datos no están protegidos por PD.

$$P(SI/está\ protegido\ por\ PD) = (1 - P1) * P2 \quad (2)$$

siendo (2) la probabilidad de obtener la respuesta SI dado que los datos están protegidos por PD.

A partir de la ecuación (1) y (2), la probabilidad de estar protegido por privacidad diferencial vendrá dada por:

$$\mathrm{P(Estar\ protegido\ por\ PD) =}$$

$$= \frac{\mathrm{P(SI) - P(SI/no\ está\ protegido\ por\ PD)}}{\mathrm{P(SI/está\ protegido\ por\ PD) - P(SI/no\ está\ protegido\ por\ PD)}} =$$

$$= \frac{\mathrm{P(SI) - [\,P1 + (1 - P1) * P2]}}{\mathrm{(1 - P1) * P2 - [\,P1 + (1 - P1) * P2]}} =$$

$$= \frac{\mathrm{P(SI) - [\,P1 + (1 - P1) * P2]}}{\mathrm{P2 - P1P2 - P1 - P2 + P1P2}} =$$

$$= \frac{-\mathrm{P1}}{-\mathrm{P1}} + \frac{\mathrm{P(SI)} - (1 - \mathrm{P1}) * \mathrm{P2}]}{-\mathrm{P1}} =$$

$$= 1 - \frac{[\,\mathrm{P(SI)} - (1 - \mathrm{P1}) * \mathrm{P2}]}{\mathrm{P1}} \qquad (3)$$

En base a la ecuación (3) puede interpretarse que mediante la aplicación de privacidad diferencial a los datos es posible obtener una reducción del riesgo de identificación de la verdadera respuesta que vendrá dado por cierto valor de probabilidad de ocurrencia. Para ello será necesario implementar algún mecanismo probabilístico para la aplicación del ruido aleatorio y que determine una distribución de probabilidad de ocurrencia. Este punto será especificado con mayor detalle en el siguiente apartado.

Aplicar esta metodología a los datos arrojará como resultado una nueva base conformada por lo datos distorsionados. A diferencia de la original contendrá datos modificados como consecuencia de haber aplicado privacidad diferencial. De este modo, cuando un usuario realice consultas no podrá notar la diferencia entre ambos en función de los resultados agregados obtenidos, ya que la distribución probabilística no cambiará sustancialmente. La aproximación a este resultado requiere de una modelización matemática, la cual se expone en el apartado 3.3. Pero para ello previamente resulta necesario introducir tres conceptos: función aleatoria, distribución de probabilidad de Laplace y el mecanismo de Laplace.

3.2 Conceptos preliminares

Antes de poder presentar el modelo de Privacidad Diferencial, resulta necesario especificar tres conceptos. En primer lugar, el significado de una función o algoritmo aleatorio. Luego, cual es el comportamiento de la distribución de Laplace. Finalmente, y en base a esta última, que es el mecanismo de Laplace.

3.2.1 Función o algoritmo aleatorio

El esquema de obtención de respuesta aleatoria presentado en el apartado anterior puede ser modelizado mediante un algoritmo aleatorio. Un algoritmo es una secuencia instrucciones elementales ejecutables, generalmente destinadas a lograr un propósito específico (Erickson, 1999). En particular, los algoritmos aleatorios son un caso especial de algoritmos estocásticos.

Un algoritmo estocástico puede ser visto como un algoritmo parcialmente controlado por un proceso aleatorio (Hromkovič, 2004). Por ejemplo, el control puede venir dado por el resultado obtenido del lanzamiento de una moneda. Este resultado será utilizado para decidir de qué manera el algoritmo continuará su proceso. La calidad de un algoritmo estocástico se mide generalmente a través del tiempo de procesamiento y del grado de fiabilidad sobre el resultado obtenido. Este último puede interpretarse como la probabilidad de obtener el resultado correcto, es decir, obtener P1 en el esquema de respuesta aleatoria planteado en el apartado 3.1. A continuación, se ejemplifica el funcionamiento de este tipo de algoritmos.

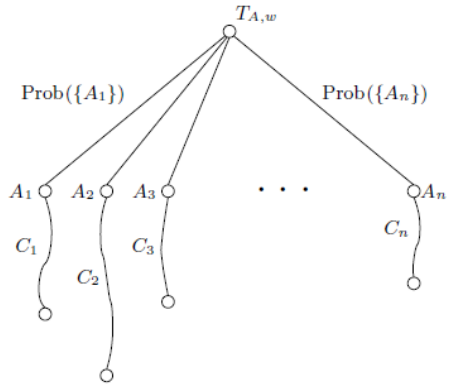
Siguiendo a Hromkovič (2004), considérese un algoritmo A como una distribución de probabilidad sobre un conjunto finito $\{\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_n}\}$ de estrategias deterministas. Para cualquier *input* *w*, A elige un $\boldsymbol{A_i}$ al azar para operar sobre *w*, es decir, cada *imput*. De esta manera habrá n cálculos (o en términos computacionales, n corridas) de A en todo el rango $\{\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_n}\}$. En consecuencia, es posible modelar el proceso de acción de A como un espacio de probabilidad siendo aquel conjunto que contiene todos los resultados posibles. Este viene dado por:

$$(S_{A,w}, Prob)$$

donde $\boldsymbol{S_{A,w}}$=$\{\boldsymbol{A_1}, \boldsymbol{A_2}, \dots, \boldsymbol{A_n}\}$ y $\boldsymbol{Prob}$ es la distribución de probabilidad en todo el rango de $\boldsymbol{S_{A,w}}$.

Ahora bien, considérese a $\boldsymbol{Prob}$ como $\boldsymbol{S_{A,w}}$=$\{\boldsymbol{C_1}, \boldsymbol{C_2}, \dots, \boldsymbol{C_n}\}$, donde $\boldsymbol{C_i}$ es cada cálculo realizado por A sobre *w*. Esto es, A elige de manera determinista cada paso i a seguir con una cierta probabilidad de ocurrencia $\boldsymbol{Prob}(\{\boldsymbol{A_i}\})$. Esquemáticamente puede ser representado del siguiente modo:

Figura 0: esquema de decisión del algoritmo estocástico



Fuente: tomado de Hromkovič (2004)

En la figura 0, $\boldsymbol{T_{A,w}}$ es el tiempo de duración del cálculo de $\boldsymbol{C_i}$. De este modo, la medición de la eficiencia del tiempo procesamiento del algoritmo vendrá dada por el máximo valor esperado del cálculo de A por cada *imput* w. Esto es:

$$\max\{ET_A(w)\} = \max\{\sum_{i=1}^{n} Prob(\{C_i\}).Time(C_i)\}$$

donde $\boldsymbol{ET_A(w)}$ es el valor esperado de cada A por cada *imput* $\boldsymbol{w}$, $\boldsymbol{Prob}(\{\boldsymbol{C_i}\})$ es la distribución de probabilidad de ocurrencia asociada a cada corrida $\boldsymbol{C_i}$ y $\boldsymbol{Time}(\boldsymbol{C_i})$ es el tiempo que tarda el algoritmo para ejecutar cada corrida $\boldsymbol{C_i}$. De esta manera, la calidad en términos de eficiencia del algoritmo vendrá dada por el tiempo máximo que tarda en ejecutar cada corrida $(\boldsymbol{C_i})$.

Para medir la fiabilidad del resultado devuelto de A sobre *w*, considérese la variable aleatoria X: $\boldsymbol{S_{A,w}} \rightarrow \{\boldsymbol{0}, \boldsymbol{1}\}$ definida como:

$$X(C_i)^{37} = \begin{cases} 1\ si\ C_i \text{ devuleve la respuesta correcta sobre } w \\ 0\ si\ C_i \text{ devuleve la respuesta incorrecta sobre } w \end{cases}$$

De esta manera, para todo $\boldsymbol{i} = \boldsymbol{1}, \ldots, \boldsymbol{n}$, el valor esperado de X es la probabilidad de éxito de A sobre *w.* Esto es:

$$\mathrm{E(X)} = \sum_{i=1}^{n} X(C_i) * Prob(\{C_i\}) =$$

$$= \sum_{i=1}^{n} 1 * Prob(\{C_i\}) + \sum_{i=1}^{n} 0 * Prob(\{C_i\}) =$$

$$= \sum_{i=1}^{n} Prob(\{C_i\})$$

siendo $\boldsymbol{Prob}(\{\boldsymbol{C_i}\})$ la probabilidad de que A devuelva la respuesta correcta, es decir, cuando X toma el valor 1.

De este modo, la representación de la aleatoriedad para la obtención de una respuesta puede modelarse a través de un algoritmo estocástico sobre el cual es posible determinar su eficiencia y confiabilidad de resultado.

[37] Esta representación emula el azar por ejemplo del lanzamiento de una moneda que solo tiene dos resultados posibles (1, si sale cara, 0 si sale cruz).

3.2.2 Distribución de Laplace

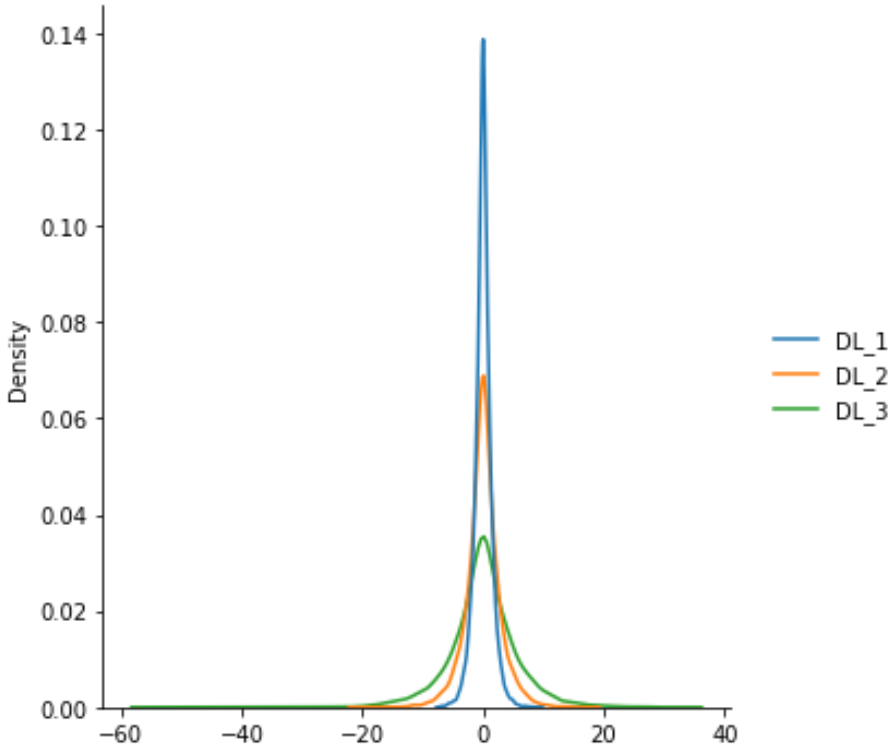
La distribución de Laplace es una distribución de probabilidad continua que lleva el nombre de Pierre-Simon Laplace. Su función de densidad de probabilidad (Dwork y Roth, 2014) está dada por la siguiente ecuación:

$$f(x) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \qquad (4)$$

donde b es el parámetro de escala, y μ es el parámetro de posición en la distribución.

Para visualizar el comportamiento de la función de densidad de probabilidad de Laplace, se crea una función en *Python* (ver código en Apéndice – sección "Apéndice A1 Distribución de Laplace") que simula el comportamiento de la función de densidad considerando que μ=0 y diferentes valores del parámetro b sobre un rango de 1000 valores. Gráficamente es posible observar lo siguiente:

Gráfico 0: Distribución de Laplace



Fuente: elaboración propia con *Python*

donde, la curva DL_1 posee un valor del parámetro b igual a 1, mientras que en el caso de la curva DL_2 toma el valor de 2 y en el caso de la cursa la curva DL_3 toma el valor de 4.

A partir del gráfico 0, se puede observar que a medida que aumenta el parámetro de escala (b), la distribución se vuelve más aplanada. Esto implica que se obtienen valores de probabilidad bajos. Tal comportamiento tendrá como consecuencia la necesidad de aumentar la distorsión en los datos para obtener mayor protección de forma tal que el dato no permita realizar una identificación del individuo. Este punto será explicitado en el apartado 3.3 con el desarrollo del modelo de Privacidad Diferencial. Para esto resulta necesario en primer lugar presentar un mecanismo que permite distorsionar datos utilizando la distribución presentada en este apartado.

3.2.3 El mecanismo de Laplace

El mecanismo de Laplace es un método estadístico por medio del cual puede obtenerse distorsión en los datos (Near y Abuah, 2021), que puede ser definido del siguiente modo. Dada una función $g(X)$ que devuelve un número (siendo una función determinística ya que se trata de por ejemplo una consulta realizada sobre una base de datos), se define una función aleatoria $F(X)$ como:

$$\boldsymbol{F(X) = g(X) + Lap\left(\frac{s}{\varepsilon}\right)} \qquad (\mathbf{5})$$

donde s es la sensibilidad de $g(X)$, y $Lap(\frac{S}{\varepsilon})$ representa una muestra de la distribución de Laplace con parámetros μ=0 y b=$\frac{S}{\varepsilon}$, cuyo comportamiento pudo observarse en la ejemplificación realizada en el apartado 3.2.2.

En la ecuación (5), puede observarse que a un resultado dado por $g(X)$ se le adiciona un valor que viene dado por $Lap(\frac{S}{\varepsilon})$ arrojando como resultado un nuevo valor $F(X)$. Este nuevo valor se encuentra distorsionado ya que al valor real (u original) se le agrega un valor aleatorio originado por de la distribución de Laplace $Lap(\frac{S}{\varepsilon})$. Además, la sensibilidad definida en términos globales de $g(X)$, viene a representar el cambio en el resultado obtenido (F(X)) cuando el valor de entrada cambia en una unidad.

En el contexto de consultas que se realizan a una base de datos, la sensibilidad global se define como la máxima diferencia entre los resultados obtenidos de consultar dos bases que tan solo difieren en un registro (Near y Abuah, 2021):

$$GS(g) = \max_{x,x':d(x,x')\leq 1} | g(x) - g(x')| \quad (6)$$

donde en la ecuación (6), $d(x, x')$ es la distancia entre dos bases de datos que tan solo difieren en un registro, es decir, que como máximo la distancia podrá valer 1.

De esta manera, al aplicar este mecanismo, se distorsiona el resultado obtenido como respuesta (los datos) al realizar una consulta sobre una base de datos. En consecuencia, la sensibilidad de la distorsión vendrá dada por la cantidad de elementos contenidos en la consulta que se realiza. Si por ejemplo una consulta cuenta el número de filas de la base de datos, y se modifica exactamente una fila de esta base, entonces el resultado de la consulta puede distorsionarse en un máximo de 1.

3.3 El modelo de Privacidad Diferencial

A partir de haber presentado los conceptos preliminares en el apartado anterior, a continuación, se presenta el modelo de privacidad diferencial. Formalmente, dado dos bases de datos ($\boldsymbol{D_1}, \boldsymbol{D_2}$), que difieren como máximo en una fila, pero una está incluida en la otra, Cynthia Dwork (2008) define a la privacidad diferencial como una función aleatoria K que permite distorsionar datos. Esto viene formalizado por:

$$\boldsymbol{Pr}[\mathrm{K}(\boldsymbol{D_1}) \in \boldsymbol{S}] \leq \boldsymbol{exp}(\boldsymbol{\varepsilon}) * \boldsymbol{Pr}[\mathrm{K}(\boldsymbol{D_2}) \in \boldsymbol{S}] \quad (7)$$

donde S está incluido en todo el dominio de K y $\boldsymbol{\varepsilon}$ es un valor que refleja la pérdida por distorsionar los datos.

En la ecuación (7) se puede observar que la probabilidad de obtener una respuesta (resultado) de $\boldsymbol{D_1}$ difiere de la probabilidad de obtener la misma respuesta de $\boldsymbol{D_2}$ al aplicar la función K. Dicha diferencia es consecuencia de haber incorporado ruido aleatorio a través de esta función K sobre los datos de $\boldsymbol{D_2}$ multiplicándola por $\boldsymbol{exp}(\boldsymbol{\varepsilon})$. A su vez, la pérdida que se produce por dicha distorsión viene dada por el valor de $\boldsymbol{\varepsilon}$. Si $\boldsymbol{\varepsilon = 0}$, entonces no hay distorsión creada – ya que $\boldsymbol{exp}(\boldsymbol{0}) = \boldsymbol{1}$– y por lo tanto los datos de $\boldsymbol{D_2}$ son iguales a los de $\boldsymbol{D_1}$ por estar los primeros contenidos en los de esta última. Si en cambio $\boldsymbol{\varepsilon > 0}$ existirá una diferencia significativa entre las probabilidades de respuesta sobre los datos de una y otra base. De este modo, el valor $\boldsymbol{\varepsilon}$ viene a representar el nivel de pérdida (o de distorsión respecto del verdadero valor del dato) como evaluación de la protección de los datos. Cuanto mayor sea el valor de $\boldsymbol{\varepsilon}$, significará que habrá una mayor protección de los datos y por tanto se brindará más garantía de privacidad ante la imposibilidad de identificar a un individuo por encontrarse los datos distorsionados.

Pero resulta necesario poder determinar cuánto ruido aleatorio se debe incorporar en los datos de forma tal de buscar un equilibrio entre la pérdida en la protección de los datos (privacidad) y la utilidad futura del dato distorsionado. Será necesario medir la sensibilidad del peso de los datos de un individuo en los cálculos que se realizan (Dwork, 2008). La medición de la sensibilidad en una base de datos en el contexto de este modelo puede ser definida del siguiente modo. Sea *g* una función que devuelve un valor $\boldsymbol{R^k}$ (donde $\boldsymbol{R^k}$ es la obtención de la respuesta verdadera, es decir, P1 en el esquema planteado en el apartado 3.1). Se define *g*: $\mathsf{D} \rightarrow \boldsymbol{R^k}$ con k ϵ [0;1]. Luego, la sensibilidad de un dato de un individuo ($\Delta \boldsymbol{g}$) vendrá dada por:

$$\Delta \boldsymbol{g} = \max_{\boldsymbol{D_1}, \boldsymbol{D_2}} \| \boldsymbol{g}(\boldsymbol{D_1}) - \boldsymbol{g}(\boldsymbol{D_2}) \| \quad (8)$$

Si en la ecuación (8), $\boldsymbol{k} = \boldsymbol{1}$ entonces *g* es la diferencia máxima entre los valores que puede tomarse en dos bases de datos que tan solo difieren en una sola fila, como se explicó en el apartado 3.2.3. De este modo, definida la sensibilidad y considerando la ecuación (7), esto viene representado a través del valor $\boldsymbol{\varepsilon}$. Cuanto menor sea el valor de $\boldsymbol{\varepsilon}$, se deberá incorporar más ruido aleatorio para garantizar un incremento en la protección de los datos.

Llegado a este punto es importante tener en cuenta que a medida que se aumenta la cantidad de consultas sobre la base de datos para incorporar ruido aleatorio se debilitará la protección de los datos. Esto es, dada la sensibilidad de $\boldsymbol{g}(\boldsymbol{X})$, la aleatoriedad comienza a perder fuerza (Dwork, 2008). Un mecanismo para sortear esta dificultad consiste en utilizar el mecanismo de Laplace para la generación de ruido aleatorio presentado en el apartado 3.2.3.

A partir del mecanismo de Laplace definido, Cynthia Dwork (2008) propone como método para la aplicación de ruido aleatorio a datos con el fin de alcanzar una cierta protección de estos a lo siguiente:

$$\boldsymbol{g}(\boldsymbol{X}) \; + \; [\boldsymbol{Lap}(\Delta \boldsymbol{g}/\boldsymbol{\varepsilon})]^{\boldsymbol{k}} \quad (9)$$

donde el parámetro de escala de la distribución de Laplace presentada en el apartado 3.2.2 viene dado por $\frac{\Delta \boldsymbol{g}}{\boldsymbol{\varepsilon}}$.

En la ecuación (9) es posible notar que la incorporación de ruido aleatorio mediante la función de Laplace independiza de la cantidad de los $\boldsymbol{k}$ componentes de $\boldsymbol{g}(\boldsymbol{X})$. Es decir, la forma de incorporar distorsión en la base de datos ya no depende de la secuencialidad de

consultas que se haga sobre la misma sino del valor de $\boldsymbol{\varepsilon}$. A medida que $\boldsymbol{\varepsilon}$ disminuye, el ruido esperado a incorporar será mayor – la curva de la distribución de Laplace será más aplanada en términos del comportamiento presentado en el apartado 3.2.2 –. De esta manera, se observará que la protección de los datos queda garantizada a partir de que la privacidad diferencial solo depende de la sensibilidad de $\boldsymbol{g}(\boldsymbol{X})$ y del parámetro $\boldsymbol{\varepsilon}$.

Desde sus fundamentos, podría decirse que el modelo se presenta como un método cuantitativo que permite brindar cierta garantía de privacidad. Al permitir distorsionar datos personales sin que se invalide su posterior utilización, podría resultar en una técnica adecuada para evitar la identificación de individuos. La Agencia Española de Protección de Datos mencionada en el capítulo 2, la ha adoptado como metodología recomendada. También grandes empresas privadas como Google y Apple la utilizan[38]. Ante este escenario prometedor, a continuación, se realiza su aplicación para evaluar su eficacia.

3.4 La Privacidad Diferencial en acción

Ante la tentadora posibilidad de proteger datos mediante la aplicación de un método cuantitativo como el presentado en el apartado anterior, realizar una aplicación del método de Privacidad Diferencial resulta adecuado. En este apartado se lleva a cabo su aplicación sobre un conjunto de datos personales y se realiza la evaluación de su efectividad en términos de protección de datos personales para evitar la identificación de individuos con estos datos. A este fin, en el siguiente apartado se comienza por especificar acerca de la base datos utilizada, para luego continuar con el proceso de identificación de individuos. Detectado el riesgo de violación de privacidad en términos de identificación de individuos, finalmente se lleva a cabo la aplicación de Privacidad Diferencial mediante la utilización del lenguaje de programación *Python*[39] en el entorno *Google Colaboratory*[40] y su posterior evaluación de resultados.

3.4.1 Obtención de datos y construcción de una base de datos personales

Para llevar adelante la ejemplificación de la aplicación de la metodología de la privacidad diferencial, resulta necesario contar con una base de datos personales. A este fin, se parte de un listado de identificador único de individuos argentinos conformado por el CUIT (Clave Única de Identificación Tributaria según la Administración Federal de Ingresos Públicos)

[38] Telefónica Tech https://telefonicatech.com/blog/privacidad-diferencial-google-apple-la-usan-con-tus-datos
[39] Acerca de *Python* https://www.python.org/
[40] Acerca de *Google Colaboratory* https://colab.research.google.com/?utm_source=scs-index

publicados en el portal de datos abiertos del Gobierno de la Ciudad de Buenos Aires[41]. Este consiste en un padrón de contribuyentes de alto riesgo, responsables del Impuesto sobre los Ingresos Brutos (IIBB) para el primer trimestre de 2015. Consta de 487514 registros y una serie de atributos, entre ellos el CUIT[42], la razón social y el tipo de contribuyente inscripto.

El CUIT se conforma de 11 dígitos. Los dos primeros se corresponden a la identificación del género (masculino o femenino) según normativa vigente hasta junio de 2021[43]. Los ocho siguientes se corresponden al Documento Nacional de Identidad (DNI) y el último es un dígito verificador. Por tanto, cada individuo posee un número de CUIT único. De este modo, el CUIT, cumple con ser un identificar único de cada individuo de la República Argentina.

A partir del padrón de CUIT, se define una muestra. Inicialmente se cuenta con 266.564 números de CUIT de ciudadanos argentinos. Sobre este conjunto identificaron los casos cuyo CUIT comienzan con número 27 y con número 20. De esta manera, se consigue obtener individuos según género (femenino en el primer caso, masculino en el segundo caso). A partir de esto se crea un nuevo atributo que es el género. La razón social es el nombre y apellido del individuo, obteniéndose un nuevo atributo identificador. Finalmente, la muestra queda conformada con 172161 CUIT pertenecientes a individuos de género masculino y 94403 CUIT pertenecientes a individuos de género femenino.

Con el objetivo de conformar una base de datos personales que contenga otros atributos personales que expongan aún más la sensibilidad de la información, se seleccionan 1000 números de CUIT al azar. La mitad de estos corresponden al género femenino y la mitad restante al género masculino. A partir de ellos, es posible obtener datos personales consultando distintos servicios *web* del Estado argentino. Si bien se lleva a cabo esta tarea a los fines de mostrar una posterior aplicación, no deja de exponer la consecuencia que implica la publicación de libre acceso de un identificador único como el CUIT. Cuanto menos se está violando el derecho a la privacidad del individuo al facilitarse el acceso a obtener más información de este sin contar con su consentimiento. Pero, además, expone el riesgo al cual es expuesto el ciudadano. Si la obtención de información cae en manos equivocadas, podría causarle daños irreparables.

[41] Portal de datos abiertos BAData. Ver https://data.buenosaires.gob.ar/dataset/
[42] El CUIT es un código que identifica de manera unívoca a trabajadores autónomos, comercios y empresas.
[43] El 8 de junio de 2021 se sanciona la normativa 5007/2021 a partir de la cual los primeros dos dígitos ya no identificaran a los individuos según sexo binario.
https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-5007-2021-350854/texto

En primer lugar, se obtiene la constancia de inscripción al Monotributo[44]. Mediante consultar gratuitamente y con libre acceso el servicio *web* de AFIP[45] es posible obtener las constancias de inscripción a esta categoría tributaria. Para ello se ingresa manualmente de manera individual cada número de CUIT y un código de seguridad por cada consulta. A continuación, se muestra una imagen del portal consultado:

Imagen 1: Servicio AFIP para la obtención de Constancia de Inscripción a Monotributo

Fuente: imagen tomada del portal web de AFIP

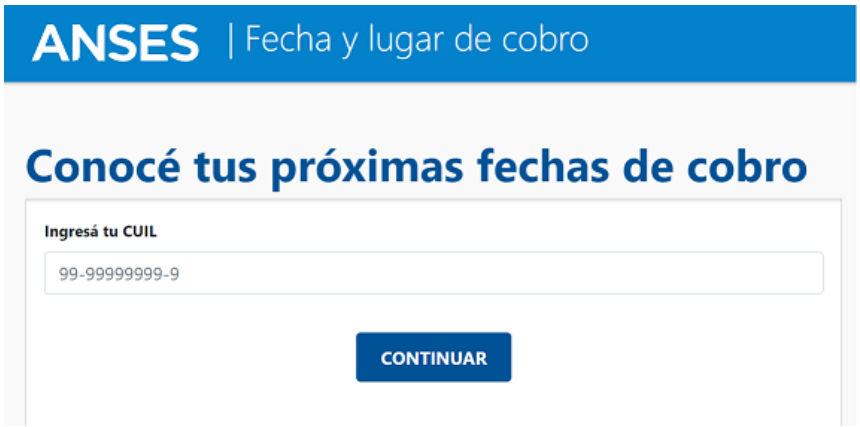
Una vez ingresados el CUIT y el código de seguridad, se obtiene en versión PDF (*Portable Document Format)* y de manera *online* la constancia de inscripción. A partir de esta se procede a copiar manualmente cada uno de los siguientes datos personales contenidos: el código postal, domicilio fiscal, jurisdicción, barrio y código de identificación de este y descripción de la actividad que realiza el contribuyente. Pero también, se verifican la razón social (nombre y apellido) y el género que se encontraban entre los datos contenidos en el listado de CUIT de inicio. Respecto de la primera, en algunos casos del listado original solo se contaba con el apellido. De este modo, también fue posible completar el nombre. Así, partiendo de un identificador único, se logra obtener 6 atributos personales más por cada contribuyente.

Finalizado lo anterior, se procede a consultar el portal *web* de ANSES[46] (La Administración Nacional de la Seguridad Social) denominado "Donde cobro". En este se debe ingresar de manera individual cada número de CUIT, sin necesidad de ingresar un código verificador.

[44] El monotributo es un régimen para pequeños contribuyentes, que unifica el pago de IVA y Ganancias con los aportes jubilatorios y la obra social (AFIP)
[45] Portal AFIP en el siguiente link https://seti.afip.gob.ar/padron-puc-constancia-internet/ConsultaConstanciaAction.do
[46] Portal "Donde cobro" de ANSES en el siguiente link https://servicioswww.anses.gob.ar/dondecobrov2/

Con tan solo un dato de identificación única, cualquier persona que lo posee puede acceder a información adicional de un ciudadano. Incluso de una manera más simple que en el caso anterior. Esto también expone que los métodos de validación que las organizaciones públicas ofrecen en este tipo de portales, cuanto menos poseen muy poca seguridad. A continuación, se muestra una imagen del portal consultado:

Imagen 2: Portal *web* "Donde cobro" de ANSES



Fuente: imagen tomada del portal web de ANSES

Al ingresar individualmente un CUIT en el mencionado portal, se logra verificar si el individuo es beneficiario social[47]. Y en los casos que lo es, se obtiene, además del nombre y apellido y CUIT, el nombre del tipo de beneficio que percibe, nombre de la organización bancaria donde percibe el cobro, así como también la dirección postal y barrio de la sucursal donde el pago es efectuado. Estos datos se copian manualmente para complementar a los anteriores, conformando un total de 10 atributos personales en la base de datos.

Finalmente, la base de datos personales es denominada "INDIVIDUOS" y fue creada en formato Excel[48]. Para su almacenamiento se utilizó Google Drive[49]. Esta contiene los datos personales para 1000 individuos. De este modo, es posible notar como la publicación de un solo identificador directo de un individuo como el CUIT, habilita la posibilidad de recolectar más información sensible sobre este. A partir de ello puede interpretarse el riesgo al cual queda expuesto el ciudadano. Pero también como su privacidad es violada en base a la

[47] Un beneficiario social de ANSES es aquel individuo que percibe alguna de las siguientes prestaciones que brinda el Estado argentino: Jubilación, Pensión, AUH, AXE, SUAF, PROGRESAR, Prestación por Desempleo. Para más información ver https://www.anses.gob.ar/sites/default/files/archivo/2021-05/Bases%20y%20condiciones%20Beneficios%20ANSES.pdf#:~:text=%E2%80%9CBENEFICIARIOS%22%3A%20son%20aquellas%20personas,(PNC)%2C%20de%20la%20Asignaci%C3%B3n
[48] Acerca de Excel https://www.microsoft.com/en-us/microsoft-365/excel
[49] Acerca de Google Drive https://www.google.com/drive/

captura de información. Como se planteó en el capítulo 2, bajo este escenario no habría respeto a la intimidad de la persona.

Finalmente, la base de datos queda conformada con los siguientes atributos como puede observarse en la tabla a continuación:

Tabla 1: Atributos que contiene la base de datos personales

Atributos	Descripción
cuit	Clave Única de Identificación Tributaria
tipo_contr_insc	Vale C o D según categoría del contirbuyente inscripto
raz_soc	Razón social (nombre y apellido del individuo inscripto)
GENERO	Vale M si es masculino y F si es femenino
CP	Código postal
Domicilio_fiscal	Domicilio fiscal
JURISDICCION	Provincia
BARRIO	Barrio
Código_actividad	Número identificatorio de actividad según nomenclador AFIP
Actividad	Nombre de la actividad que realiza según nomenclador AFIP
Beneficiario_social	Vale SI si es beneficiario social, NO en caso contrario
Nombre_beneficio_social	Nombre del beneficio social
Lugar_de_cobro	Nombre de la organización bancaria, dirección postal y barrio de la sucursal de percepción del cobro

Fuente: elaboración propia

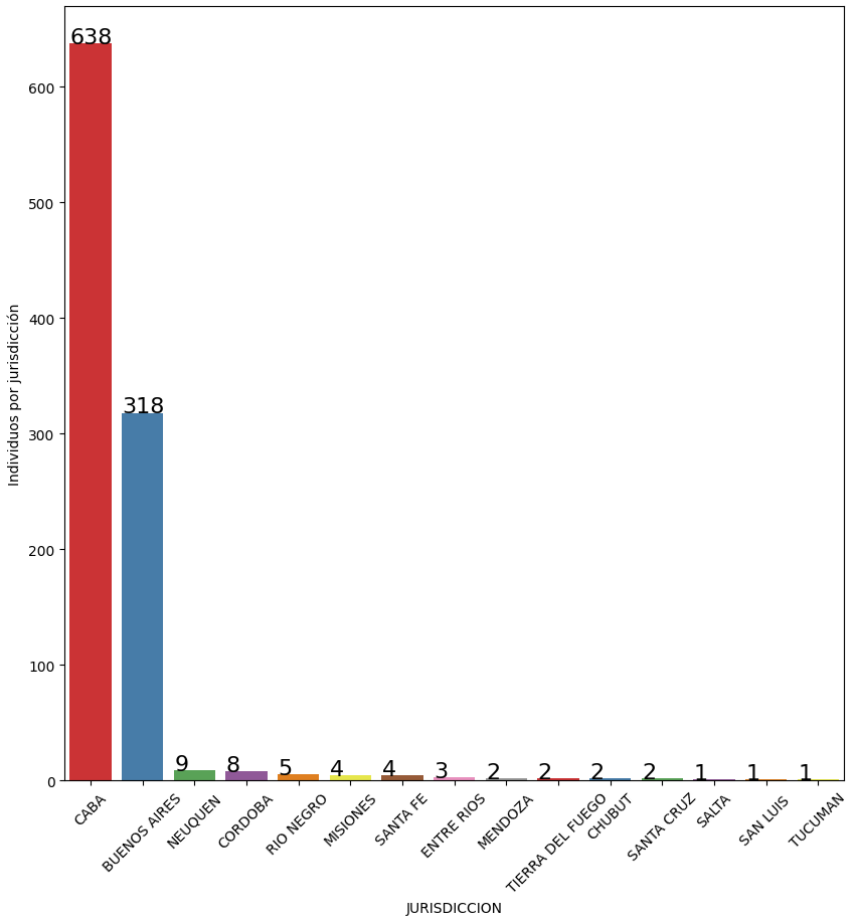
De esta manera, se logró recolectar información personal de 1000 individuos argentinos, siendo 500 de género femenino y 500 de género masculino. La información de código postal, dirección postal, jurisdicción y barrio se obtuvo de manera completa en todos los casos. Respecto de la actividad que realizan, solo en 508 se logró identificar, ya que para los restantes no figuraba en su constancia de inscripción a Monotributo. En cuanto a los datos obtenidos de ANSES, 401 resultaron ser beneficiarios sociales, obteniéndose datos completos. El 59,9% restante resultó ser no beneficiario.

3.4.2 Análisis descriptivo de la base de datos personales

A partir de la base de datos conformada para un total de 1000 ciudadanos argentinos, se realiza un análisis descriptivo inicial de las principales características observadas. Entre las características principales que presenta la base de datos personales construida se encuentra la distribución geográfica de los 1000 individuos según jurisdicción. El 63,8% (638) de los individuos tiene domicilio fiscal en CABA (Ciudad Autónoma de Buenos Aires), el 31,8% (318) en Provincia de Buenos Aires y el 4,4 % restante en diferentes provincias de Argentina.

La distribución de individuos según jurisdicción se puede observar en el gráfico a continuación:

Gráfico 1: Cantidad de individuos según jurisdicción

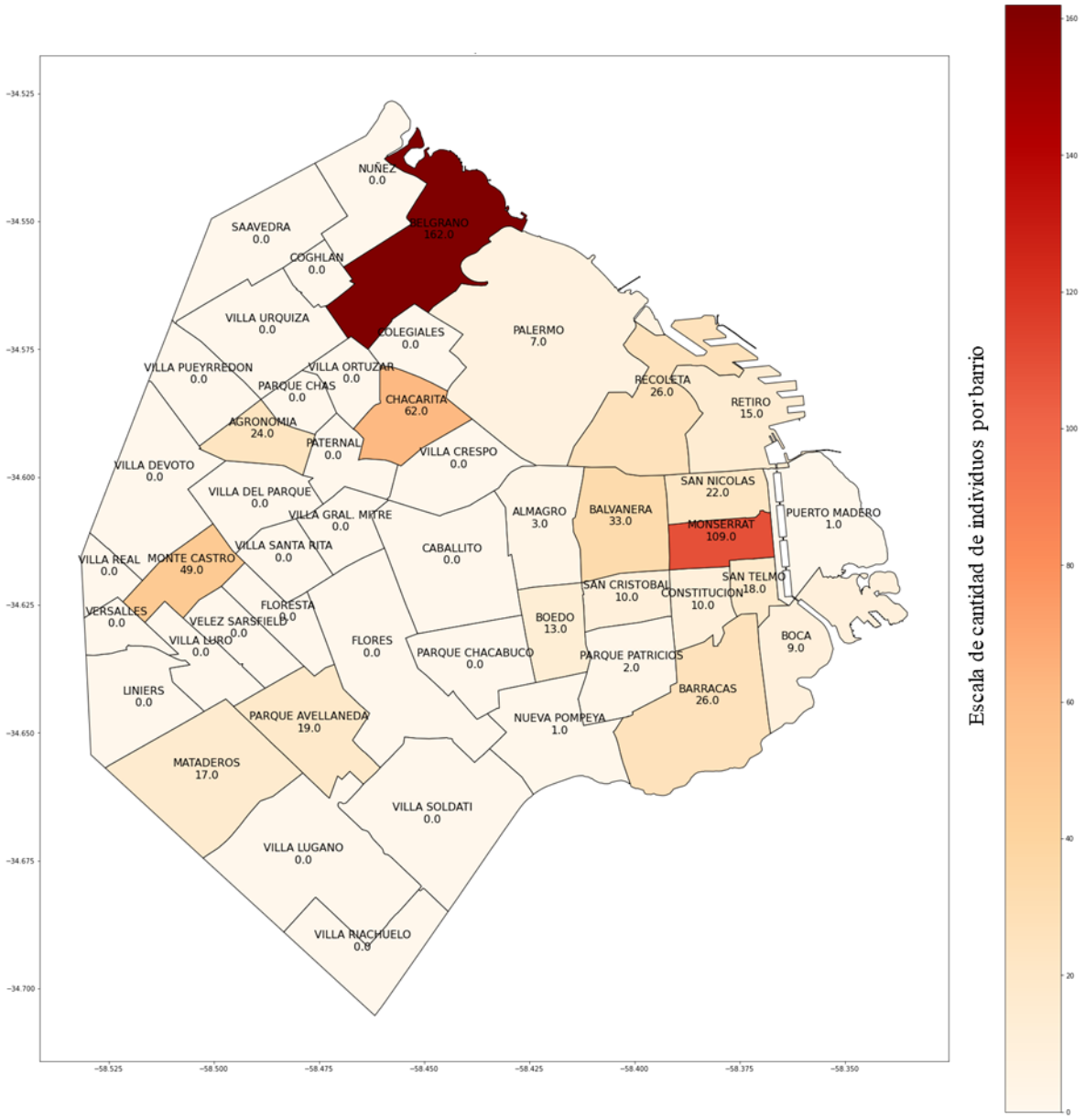


Fuente: elaboración propia con Python

En el gráfico 1, se puede observar que en las jurisdicciones del interior existe una cantidad muy baja de individuos. Particularmente, en el caso de Salta, San Luis y Tucumán hay un solo individuo por cada una. Esto muestra que con tan solo la jurisdicción un individuo puede ser identificable dentro de una base de datos.

Entre los 638 individuos de CABA, el 25,4% se encuentran en el barrio de Belgrano. Pero también es posible observar casos únicos por barrios. Para exponer de una manera visualmente sencilla la información, a continuación, es posible observar el siguiente mapa de CABA por barrio junto la cantidad de individuos:

Mapa 1: Cantidad de individuos de CABA según barrio



Fuente: elaboración propia con Python

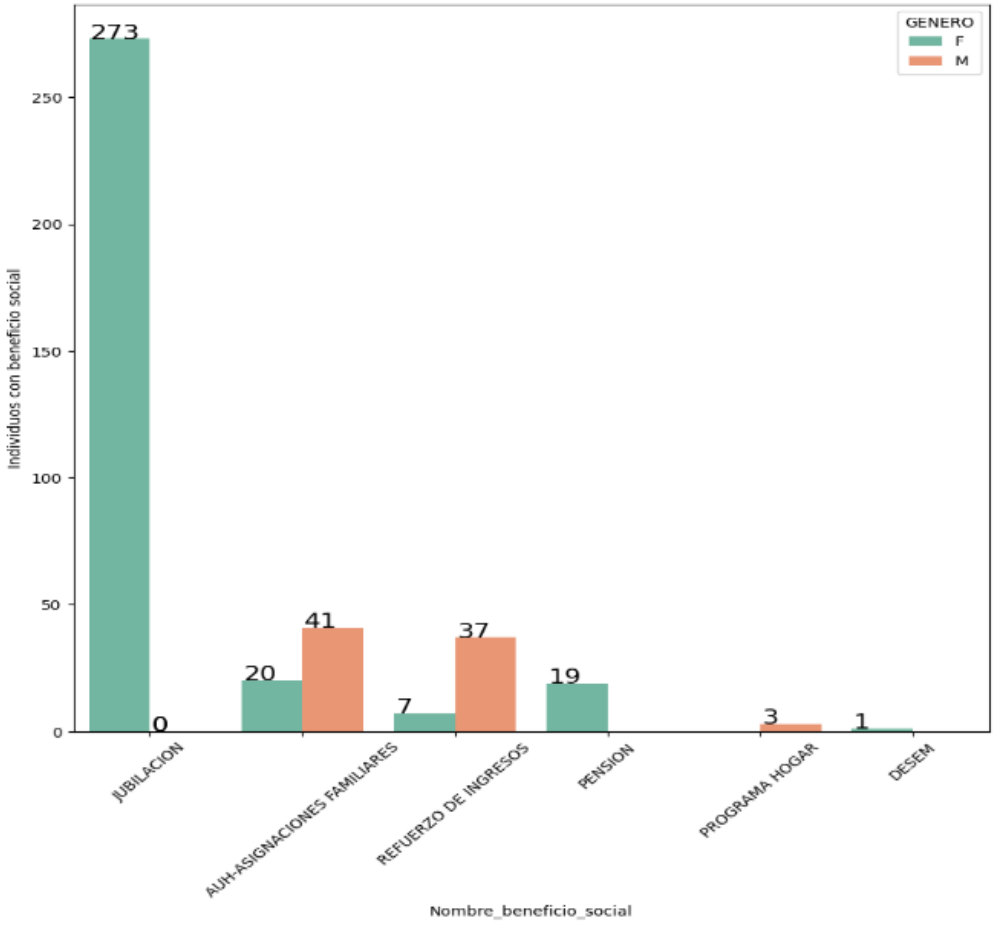
En el mapa 1 se puede observar que los barrios de Belgrano y Monserrat poseen la mayor cantidad de individuos. Pero, al mismo tiempo, se observa que en el barrio de Nueva Pompeya y Puerto Madero, existe un solo individuo en cada uno. Esto muestra que con tan solo el barrio un individuo puede ser identificable dentro de una base de datos.

Por otra parte, entre aquellos individuos (508) que pudo constatarse la actividad que ejercen en la constancia de Monotributo, se observó que la principal actividad que realizan es Servicios Personales (61 individuos). En segundo lugar, se encuentra Servicios Inmobiliarios (28 individuos) por cuenta propia y en tercer lugar Servicios Jurídicos (22 individuos). A su vez, en función de la categoría de Monotributo, es posible decir que 610 individuos poseen

categoría D y 390 categoría C. La categoría está relacionada con el monto bruto de facturación anual permitida según AFIP. En el caso de la D implica una facturación bruta anual de como máximo $ 1.934.273,04, mientras que en el caso de la C es de como máximo $ 1.557.443,75[50].

En cuanto a los datos obtenidos del portal de ANSES, 401 individuos resultaron tener algún beneficio social. De estos, el 68.08% corresponde a Jubilación mientras que el 15,2% corresponde a AUH (Asignación Universal por Hijo) y el resto a otros beneficios. Para visualizar está información, considerando también tomar el atributo género, a continuación, se presenta la distribución de quienes poseen algún beneficio social según el beneficio que perciben y su género.

Gráfico 2: Beneficiarios sociales, según tipo de beneficio y generó



Fuente: elaboración propia con Python

En el gráfico 2, se puede observar que todos los casos que poseen Jubilación son mujeres. Lo mismo sucede en el caso de Pensión. En cambio, donde existe mayor cantidad de

[50] Datos según AFIP a julio de 2022.

individuos con género masculino es en el caso de AUH, Refuerzo de Ingresos y Programa Hogar. De esta manera, de un total de 500 individuos de género masculino, solo el 16,2% posee algún beneficio social. En el caso de los 500 individuos de género femenino esta cifra alcanza al 64%.

3.4.3 Identificación directa de individuos

Conformada la base de datos personales, la primera identificación directa de individuos que surge es la realizada a través del CUIT. Al tratarse de una identificatoria única por cada individuo argentino, además permitió obtener mucha más información personal de estos consultando diferentes servicios *web* gratuitos del Estado. De aquí que no debiera publicarse el mismo, o si se publica debiera ser en forma anonimizada. Esto debido a que la información posible de obtener es accesible por cualquier persona y puede ser utilizada en perjuicio de su titular. Por esta razón, como una primera medida se decide anonimizar el CUIT, respetando un principio ético del uso de datos para procesos de investigación (Saunders et al., 2016).

Mediante la anonimización del CUIT con *Python*, se crea una nueva variable en la base datos denominada "Fake_cuit" y se procede a eliminar la original ("cuit"), manteniendo todo el resto de los atributos. Pero, aun así, surge una problemática de identificación directa de individuos según cantidad de casos por jurisdicción. En el gráfico 1 mostrado en el apartado anterior fue posible observar que, en el caso de Salta, San Luis y Tucumán, existe un solo individuo por cada una de estas jurisdicciones. Esto implica una identificación única de caso por jurisdicción. Ante esta situación, deberían eliminarse los casos – o por lo menos anonimizarse los datos completos de cada individuo involucrado– si se quiere evitar su identificación. Con esto se pretende exponer que no solo los identificadores únicos como el CUIT pueden facilitar la identificación unívoca de un individuo en una base datos. Otros atributos pueden resultar en uno de este tipo según la cantidad de casos involucrados en la base de datos.

Similar situación se presenta en el caso de los individuos de CABA por barrio. Como fue posible observar en el Mapa 1 elaborado en el apartado anterior, en los barrios de Nueva Pompeya y Puerto Madero, existe un solo individuo en cada uno. Esto implica una identificación única de caso por barrio y muestra que con tan solo este atributo un individuo puede ser identificable dentro de una base de datos. De este modo, atributos geográficos

como la jurisdicción o el barrio, requieren de una evaluación previa antes de poner libremente el dato accesible para cualquier usuario.

También se detecta que es posible identificar individuos a partir de la combinación de dos atributos. En el gráfico 2 del apartado 4.1, se observó que para el caso de género masculino y tipo de beneficio social "DESEM" existe un solo caso. De este modo, como ha sido mencionado en el apartado 2 del presente trabajo, la identificación de individuos en una base de datos personales no solo es posible en base a un identificador directo, sino también a partir de combinar atributos. Así, en una instancia inicial tan solo de descripción de una base de datos personales, queda expuesto el riesgo de identificación de individuos por falta de protección de los datos personales. Esto a su vez, expone el riesgo de violación de la privacidad.

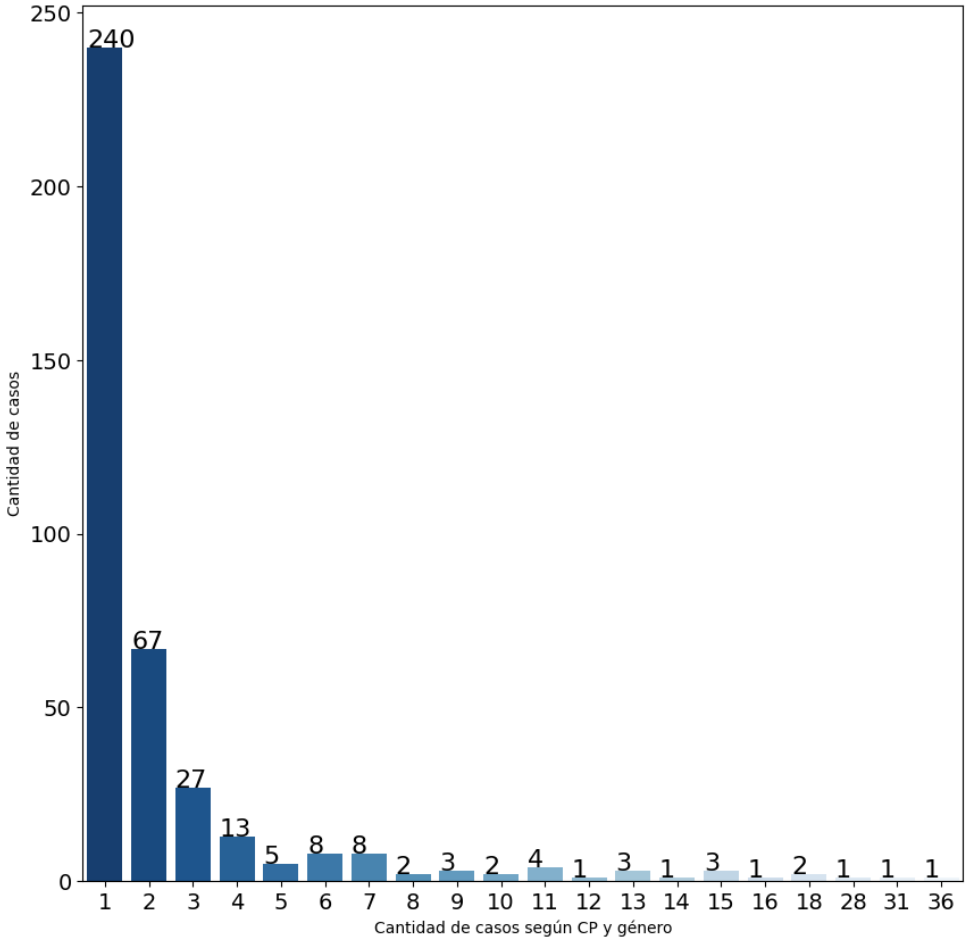
Pero también, es de destacar que para la obtención de la base de datos personales se utilizaron diversas fuentes de información. Esto expone como la combinación de datos provenientes de orígenes distintos conlleva a la identificación directa o indirecta de los ciudadanos. De aquí que el riesgo de violación de la privacidad puede escalar aún más. Esta práctica de transferencia (u acceso para la obtención) de datos entre organizaciones requiere de ser controlada de manera preventiva y no reactiva. De este modo, la combinación inicial de atributos mostrada (género con el Beneficio Social) refleja una situación similar a lo sucedido con el caso de uso indebido de datos biométricos por parte de un Ministerio de la Ciudad de Buenos Aires, mencionado en la introducción del capítulo 2. En este último caso los datos propios del Ministerio en cuestión se combinaron con los provenientes de RENAPER para la identificación de individuos.

A su vez, se pone de manifiesto que la regulación puede resultar insuficiente para evitar casos como los mencionados en el párrafo anterior. Esto se debe a que actúa una vez que el hecho sucedió como se argumentó en el capítulo 3. De aquí que no evita el perjuicio causado a los ciudadanos dado que no contempla acciones preventivas. Así, la recombinación de datos personales se convierte en un acto relativamente simple de llevarse a cabo y expone el alto riesgo que implica en términos de privacidad. En el siguiente apartado se profundiza sobre el caso de reidentificación de individuos a partir de la combinación de atributos personales.

3.4.4 Identificación indirecta de individuos

En el apartado 4.1.2 se expusieron diferentes casos de identificación directa de individuos a partir del uso de atributos personales únicos que ponen en riesgo la privacidad de los individuos. Pero también, se introdujo un nuevo caso que surge de la recombinación de atributos personales dentro de una misma base de datos. Para ejemplificar la cantidad de individuos que tienen una combinación única de atributos en la base de datos bajo análisis, a continuación, se muestra gráficamente considerando el código postal (CP) y el género.

Gráfico 3: Cantidad de individuos identificables según combinación de código postal y género



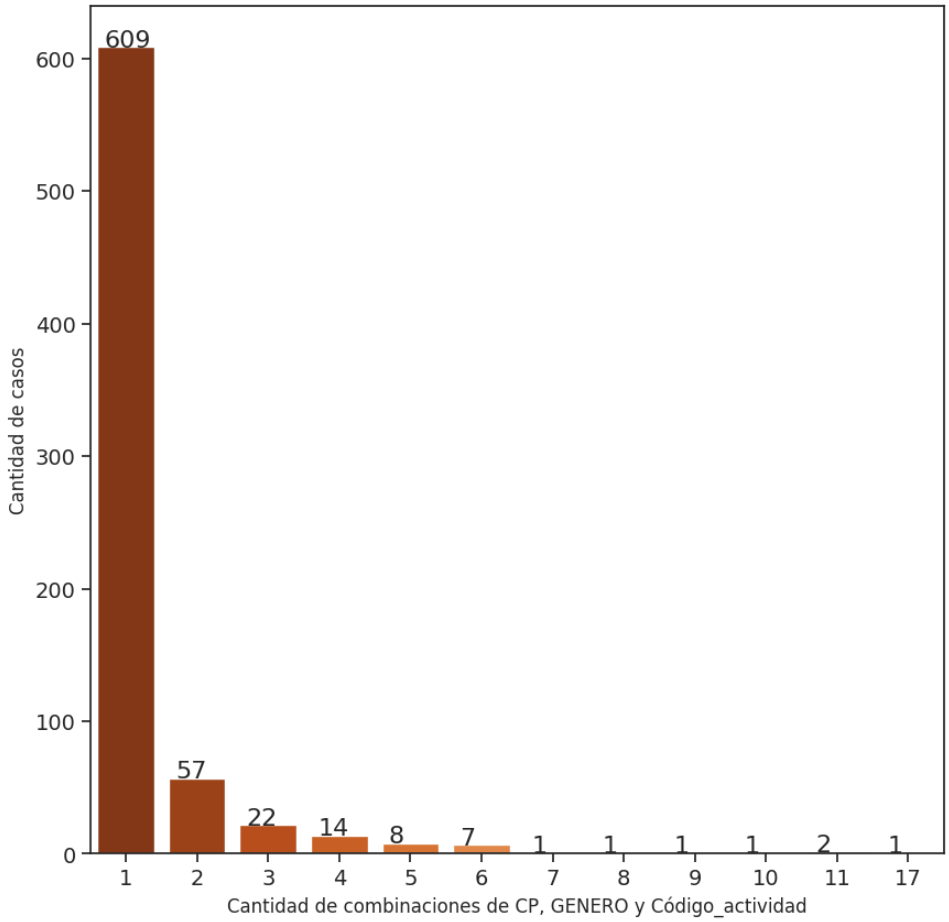
Fuente: elaboración propia con Python

El gráfico 3 muestra en el eje de abscisas la cantidad existente en la base de datos de combinaciones de código postal y género (vale 1 si es única); y en el eje de ordenadas, la cantidad de individuos. Su interpretación es como sigue: existen 240 individuos de un total de 1000 (24%) en la base de datos que poseen una combinación única (no repetida) de ambos atributos (en el eje de abscisas toma el valor 1). Mientras que existen 67 casos donde la combinación de los atributos existe dos veces (en el eje de abscisas toma el valor 2), es decir,

que involucra a un total de 134 individuos. Así, es posible interpretar los restantes casos para obtener el total de 1000 individuos. Por lo tanto, en el primer caso, la identificación del individuo es unívoca dado que la combinación es única. En el segundo caso mencionado no lo es, ya que existen dos individuos por cada combinación de atributo.

Si ahora se incorpora un atributo más, por ejemplo, el código de actividad que realiza, entonces los casos obtenidos se muestran en el gráfico a continuación.

Gráfico 4: Cantidad de individuos identificables según combinación de código postal, género y código de actividad

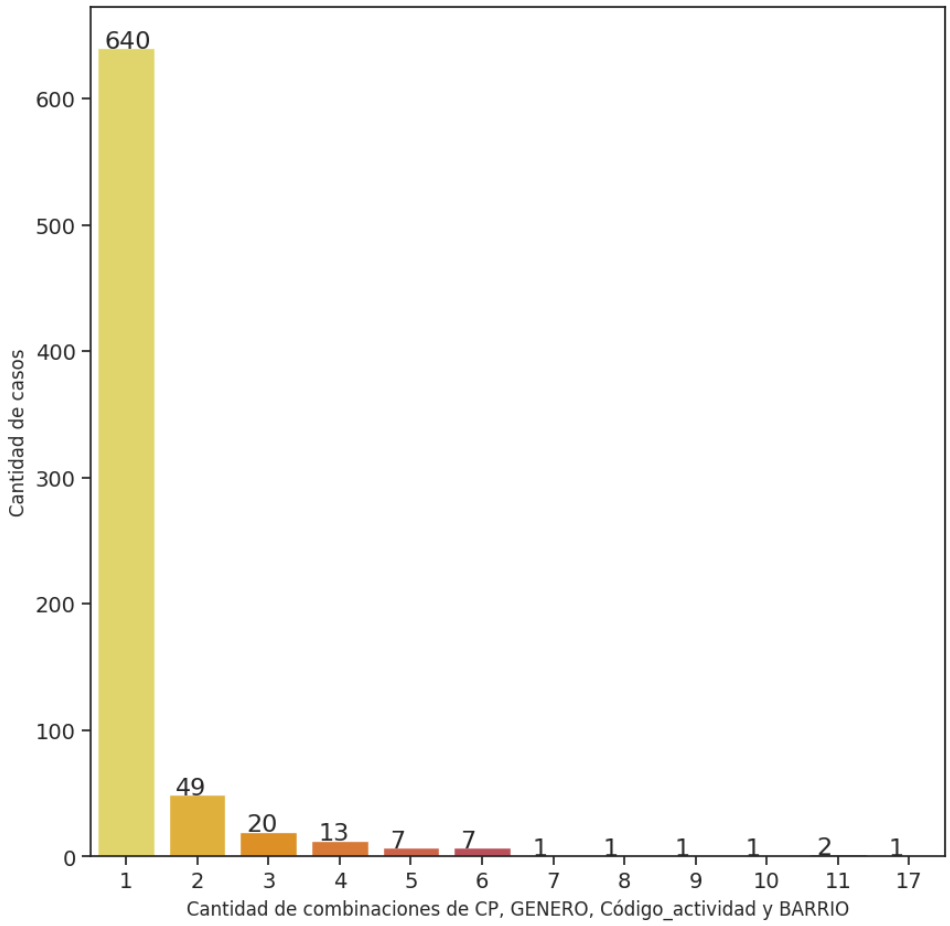


Fuente: elaboración propia con Python

A diferencia de la situación del gráfico 3, en el gráfico 4 se observa que la cantidad de individuos identificables unívocamente a partir de combinar tres atributos aumenta a 609, siendo el 60.9% del total de la base de datos. De aquí que, a medida que se aumenta la cantidad de atributos combinados en una misma base de datos, aumenta el riesgo de identificación de un individuo y, por tanto, de violación de su privacidad.

Si ahora se incorpora un atributo más, por ejemplo, el Barrio, entonces los casos obtenidos se muestran en el gráfico a continuación.

Gráfico 5: Cantidad de individuos identificables según combinación de código postal, género, código de actividad y barrio

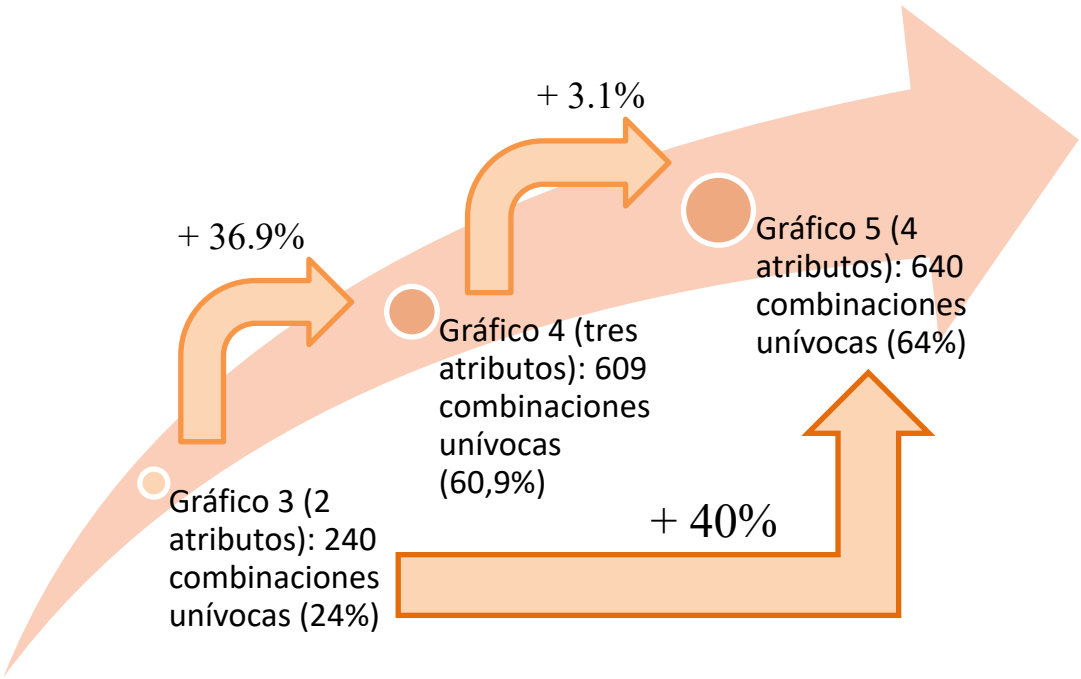


Fuente: elaboración propia con Python

En el gráfico 5 se observa que la cantidad de individuos identificables unívocamente a partir de combinar cuatro atributos aumenta a 640, siendo el 64% del total de la base de datos. De esta manera, resulta evidente que a medida que se combina una mayor cantidad de atributos, se logra aumentar los casos de identificación de individuos en la base de datos. Si se compara la cantidad de casos identificables de manera unívoca obtenidos en el gráfico 3, 4 y 5, es posible determinar un incremento del riesgo de identificación de individuos.

Para dimensionar como se incrementa el riesgo de identificación de individuos a medida que se combina una mayor cantidad de atributos, a continuación, se expone en una figura:

Figura 3: incremento del riesgo de identificación de individuos



Fuente: elaboración propia

En la figura 3, puede observarse como aumenta el riesgo de identificación de individuos a medida que se combinan más atributos personales. Para lo casos analizados previamente, puede observarse que el riesgo sufre un incremento del 40% cuando se utilizan 4 atributos combinados respecto de utilizar dos en la base de datos bajo análisis. De aquí que, la recombinación de atributos personales resulta ser una acción muy riesgosa y no tan difícil de llevarse a cabo. Ante esto, la problemática de identificación de individuos invita a preguntarse qué atributos son publicables en conjunto y cuales no cuando se cuenta con una única base de datos.

Ahora bien, es importante no perder de vista que la recombinación de atributos no necesariamente surge dentro de una misma base de datos. De combinar datos provenientes de diferentes bases, también es posible la reidentificación de individuos. Teniendo en cuenta que entre organizaciones del Estado es frecuente la práctica de transferencia de datos para el diseño de políticas públicas, resulta interesante evaluar el riesgo asociado a la identificación de individuos que puede surgir. Este caso, también expone el riesgo de la privacidad del individuo.

Para llevar adelante la demostración de este caso particular, en primera instancia se particiona en dos la base de datos “INDIVIDUOS” con el CUIT ya anonimizado. Como

resultado se obtiene, por un lado, una base (individuos_1) que contiene 'tipo_contr_insc', 'raz_soc', 'GENERO', 'CP', 'Domicilio_fiscal', 'JURISDICCION', 'BARRIO', 'Código_actividad', 'Actividad' y 'Fake_cuit'. Por el otro, una segunda base (individuos_2) con los atributos 'GENERO', 'CP', 'Beneficiario_social', 'Nombre_beneficio_social', 'Lugar_de_cobro' y 'Fake_cuit'. En cada una de ellas se mantiene el total de individuos de 1000. De esta manera se emula la posibilidad de obtener una base datos proveniente de una fuente externa y otra obtenida de los propios sistemas de información de una organización.

A continuación, si se agrupan los datos de ambas bases según código postal, género y código de actividad, en la base de datos individuos_1 se obtiene que hay 609 casos con combinación única, como se había mostrado en el gráfico 4. Luego, a modo de ejemplo, se toma un caso de estos 609, cuyo código postal es 1002, el género es femenino y el código de actividad es 702091. Si a continuación se combina esta información con el segundo conjunto de datos (individuos_2) se obtiene unívocamente un caso con toda la información. Los datos completos del individuo involucrado son[51]:

	tipo_contr_insc	raz_soc	GENERO	CP	Domicilio_fiscal	JURISDICCION	BARRIO	Código_actividad	Actividad	Fake_cuit	Beneficiario_social	Nombre_beneficio_social	Lugar_de_cobro
0	D		F	1002		CABA	MONSERRAT	702091	SERVICIOS DE ASESORAMIENTO, DIRECCIÓN Y GESTIÓ...	XNZF0716667456452 4	SI	JUBILACION	BBVA- BANCO FRANCES S.A. RECONQUISTA 00199 - ...

De esta manera se muestra como mediante la combinación de datos personales provenientes de diferentes bases es posible identificar unívocamente a los individuos. A su vez esto permite ampliar la cantidad de información sobre el individuo lo que facilita aún más su identificación definitiva. Al mismo tiempo, también queda expuesto como la anonimización de algún identificador único puede resultar insuficiente para evitar este riesgo. Ahora bien, podría pensarse en anonimizar todos los atributos. Pero si se realiza esto, se convertirían los datos en códigos sin sentido, como puede observarse en el caso del CUIT en la base de datos INDIVDUOS, y como fue mencionado en el capítulo 2, inhabilitando su uso posterior.

Por esta razón, la anonimización de datos resulta ser una técnica de uso limitado. Si se quiere conservar datos útiles para un análisis posterior, no es posible aplicarla a toda la base de datos. En consecuencia, se requiere de otras metodologías para sortear este inconveniente. En este trabajo y como fue presentado en el capítulo 3, se sugiere la utilización de la

[51] No se muestran el nombre y apellido ni dirección postal para proteger los datos del individuo involucrado.

Privacidad Diferencial mediante el mecanismo de Laplace. Para poder evaluar la efectividad de esta metodología, en el siguiente apartado se lleva a cabo su implementación y evaluación.

3.4.5 Análisis de la efectividad de la metodología de Privacidad Diferencial para la protección de datos personales

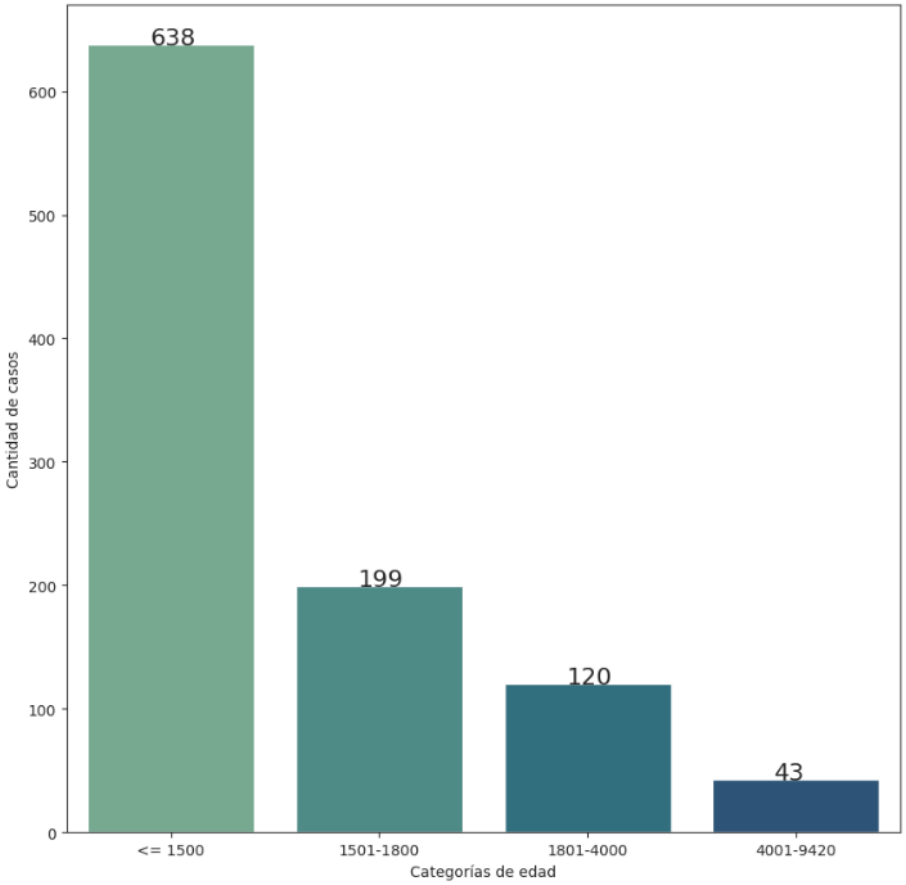
Para poder llevar a cabo la implementación de Privacidad Diferencial (PD) sobre datos personales resulta necesario que lo atributos sean de tipo numérico. Esto porque, el método, es cuantitativo como pudo observarse en la presentación realizada en el apartado 3. Una vez que se aplica generando distorsión sobre el atributo, su efectividad vendrá dada por alcanzar una cierta protección – es decir, rotura de la posibilidad de identificar individuos por haber distorsionado el dato–. Finalmente, la evaluación de su utilidad podrá derivarse de analizar si la distribución del dato distorsionado cambia o no respecto de la distribución del dato original. A su vez, podrá determinarse que nivel de privacidad (o distorsión) será necesario alcanzar realizando diferentes pruebas para distintos valores del parámetro épsilon y midiendo el error alcanzado. A partir de ello podrá determinarse el nivel adecuado de protección o privacidad a aplicar.

Con el fin de poder aplicar PD y analizar su efectividad, se selecciona el atributo código postal (CP) de la base datos individuos_2 ya que es numérico. Además, es un identificar indirecto que como se mencionó en el apartado 3.1 combinado con otros permite identificar individuos en una base de datos. De este modo, mediante la utilización de la librería OpenDP[52] con *Python* se lleva a cabo la aplicación de PD y su posterior análisis.

Antes de comenzar con la aplicación de PD, se observa que el CP en individuos_2 posee un valor mínimo de 1002 y un valor máximo de 9420. Para poder observar su distribución, se lo agrupa en ciertos rangos donde cada uno de estos contendrá a un cierto grupo de individuos. Para exponerlo de un modo simple, se realiza un gráfico de barras que es representativo de la distribución de individuos según CP. En el gráfico a continuación, puede observarse la distribución:

[52] Acerca de OpenDP https://pypi.org/project/opendp-smartnoise-core/

Gráfico 6: Cantidad de individuos según rangos de código postal



Fuente: elaboración propia con *Python*

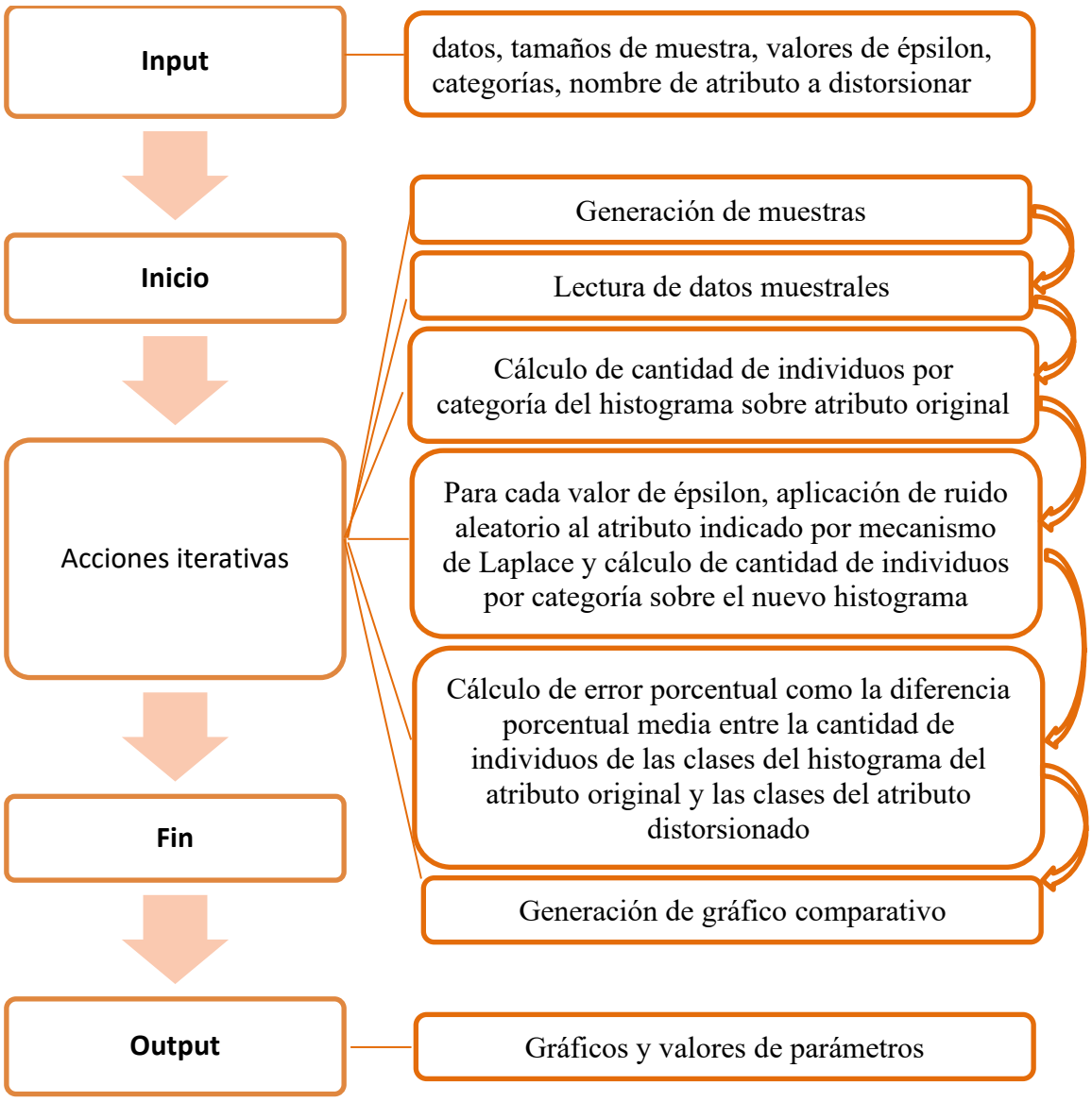
A partir del gráfico 6 puede observarse que la mayor cantidad de individuos contenidos en la base de datos individuos_2 se encuentran entre los códigos postales 1002 y 1500. Dada esta distribución original del CP recategorizado a continuación se aplica privacidad diferencial sobre este. Para llevarlo a cabo se desarrolló una función con *Python* que contiene a otras funciones ya programadas pertenecientes a la librería OpenDP.

La función desarrollada consiste en una serie de instrucciones que van desde la determinación de diferentes muestras de datos, su lectura, aplicación de la distorsión sobre el atributo indicado por el mecanismo de Laplace para diferentes valores de épsilon, cálculo del error obtenido y obtención de gráficos comparativos. Esto se lleva a cabo de manera iterativa por cada muestra y, para cada una de estas, por cada valor de épsilon, utilizando el método del histograma. El histograma permite obtener el comportamiento de la distribución de un atributo numérico a través de representar la frecuencia (absoluta o relativa) de los casos involucrados por categorías (intervalos de clase). Por lo tanto, cualquier afectación en los casos (quitar o agregar un registro en la base de datos) puede cambiar drásticamente la

distribución ya que son muy sensibles a cambios en los valores que contiene cada clase. Por esta razón es que en el contexto del modelo a aplicar resulta adecuada su representación.

De este modo, la función creada (Ver código en Apéndice – sección "Apéndice A2 Aplicación de Privacidad Diferencial y Evaluación de Resultados" – "6. Aplicación de Privacidad Diferencial") constituye un algoritmo con la siguiente estructura:

Figura 4: algoritmo de aplicación de Privacidad Diferencial



Fuente: elaboración propia

Programado con Python el algoritmo expresado en la Figura 4 se procede a aplicarla sobre el atributo CP. Para ello se considera tomar diferentes tamaños de muestra (250, 500, 750, 850, 950), distintos valores del parámetro ε (0.8, 0.5, 0.25, 0.05), y las categorías de código

postal ['<= 1500', '1501-1800', '1801-4000','4001-9420'] mostradas en el gráfico 5. A partir de ello se obtienen los resultados contenidos en la tabla 2 a continuación.

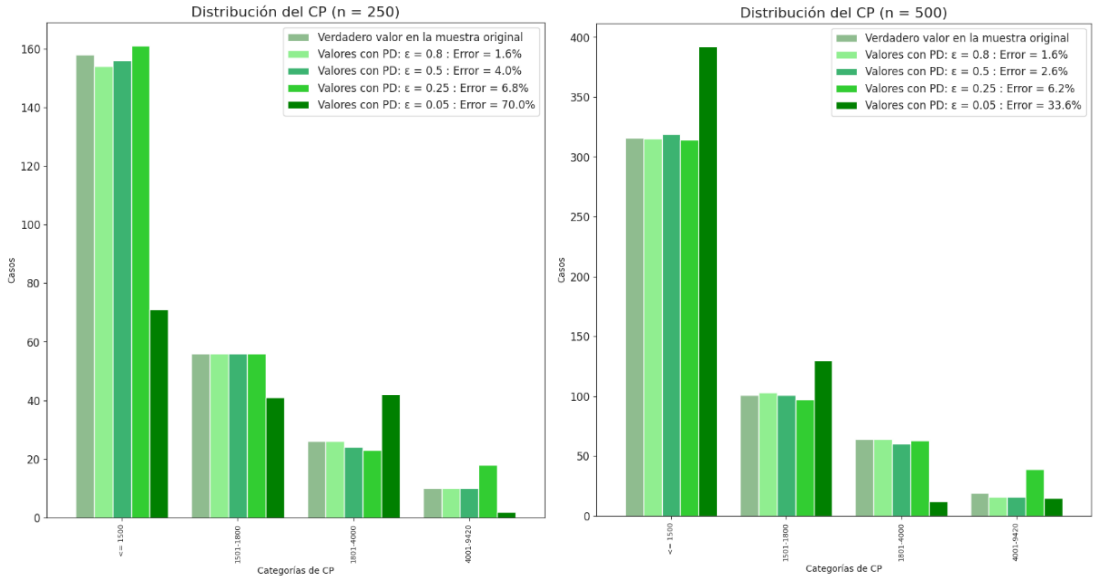
Tabla 2: Resultados de la aplicación del modelo de Privacidad Diferencial

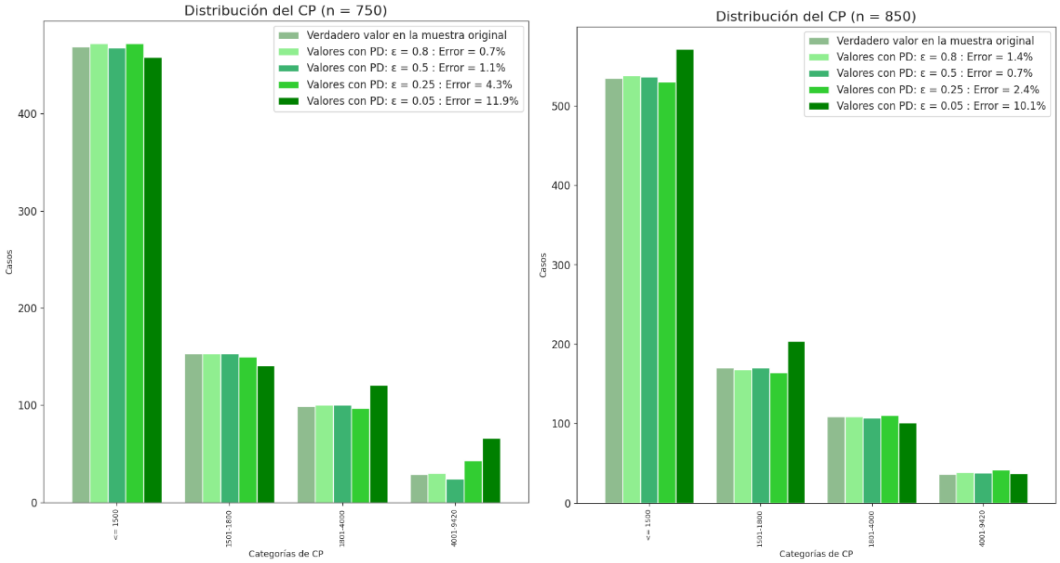
Tamaño de muestra	250	500	750	850	950
	Error obtenido				
PD con épsilon (ε=0.8)	1,6%	1,6%	0,7%	1,4%	0,5%
PD con épsilon (ε=0.5)	4,0%	2,6%	1,1%	0,7%	1,2%
PD con épsilon (ε=0.25)	6,8%	6,2%	4,3%	2,4%	1,5%
PD con épsilon (ε=0.05)	70,0%	33,6%	11,6%	10,1%	8,9%

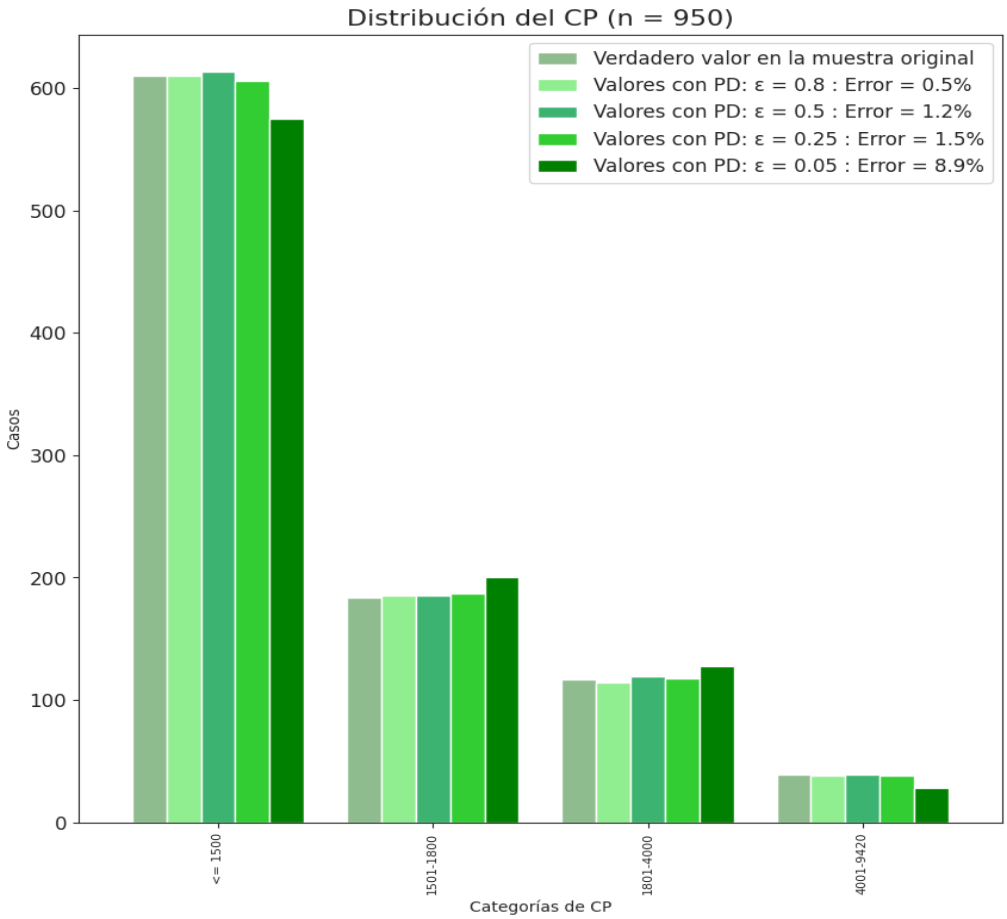
Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

A partir de los resultados mostrados en la Tabla 2, se puede concluir que, para cada tamaño de muestra, a medida que disminuye el valor de ε – es decir, a medida que se opte por reducir la pérdida de privacidad (o aumentar la distorsión)–, el error aumenta. Este resultado es esperable. Se debe a que como se definió anteriormente el error representa en términos porcentuales la variabilidad (diferencia entre verdadero valor y valor distorsionado) media que existe entre las clases de cada histograma. El algoritmo elabora un histograma con la muestra original y otro con los datos obtenido a partir de distorsionar el atributo. Pero, a su vez, es posible observar que a medida que se aumenta el tamaño de la muestra, a igual valor del parámetro ε, el error disminuye. Para visualizar estos resultados en la distribución del atributo código postal, a continuación, se muestran gráficamente los resultados:

Gráfico 7: Cantidad de individuos según rangos de código postal con y sin privacidad diferencial (para diferente valor de épsilon) y diferentes tamaños de muestra







Fuente: elaboración propia con Python

De los gráficos 7 puede observarse que la distribución del atributo CP en categorías no se ve fuertemente modificada por haber aplicado el método de privacidad diferencial. Además, a medida que se aumenta el tamaño de la muestra y para valores no tan bajos de ε, se alcanza niveles medios altos de distorsión (y, por tanto, protección) del atributo personal. Esto

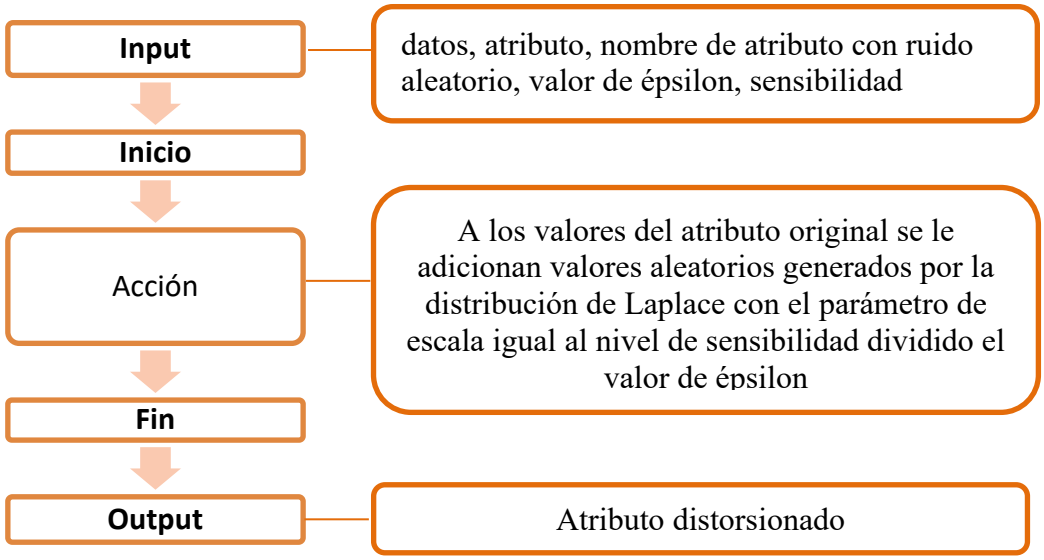
implica que el dato distorsionado no permitirá realizar la identificación de individuos ya que no se cuenta con el valor original. De aquí que el método resulta efectivo para proteger datos personales.

De este modo, considerando los resultados obtenidos, mediante la aplicación de la metodología de Privacidad Diferencial es posible obtener datos distorsionados sin que se afecte la distribución de un atributo en comparación con la original. Con ello se logra mostrar la efectividad de aplicar el método para la base utilizada. Como resultado se obtiene que su aplicación no invalida la utilidad del dato y por lo tanto podrá ser utilizado para cualquier análisis posterior. De este modo, es posible un uso responsable de datos personales al brindar garantía de privacidad a los individuos en los términos definidos en este trabajo.

3.4.6 Aplicación final de ruido aleatorio y evaluación de resultados

A partir de haber evaluado la efectividad de la metodología de Privacidad Diferencial para la protección de los datos personales contenidos en la base utilizada en el apartado previo, a continuación, se aplica sobre el atributo código postal del conjunto de datos original (INDIVIDUOS). Para realizarlo se crea una función que modela la ecuación (9) del apartado 3.3 (Ver código en Apéndice – sección "7. Implementación final de ruido aleatorio y evaluación de resultados" – sección 7.1). La estructura del algoritmo desarrollado es el siguiente:

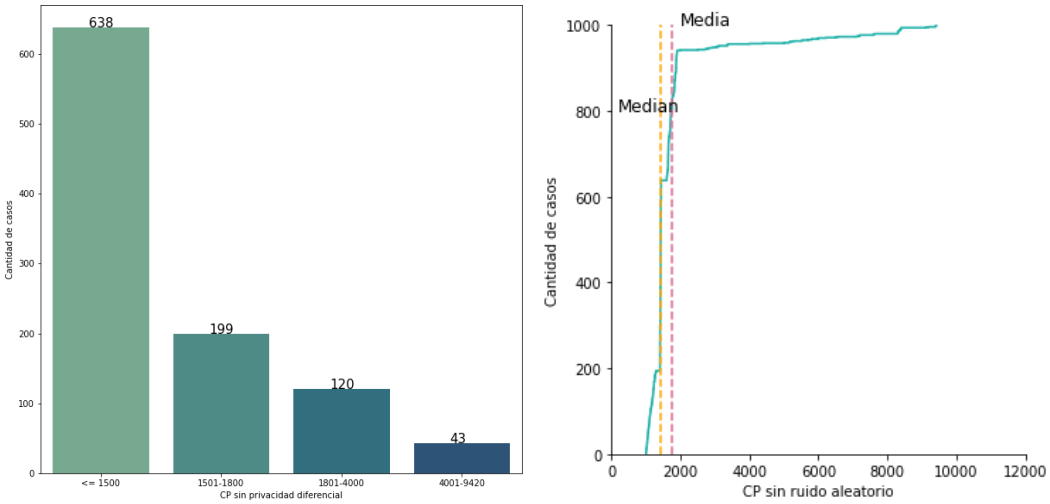
Figura 5: algoritmo de aplicación ruido aleatorio



Fuente: elaboración propia

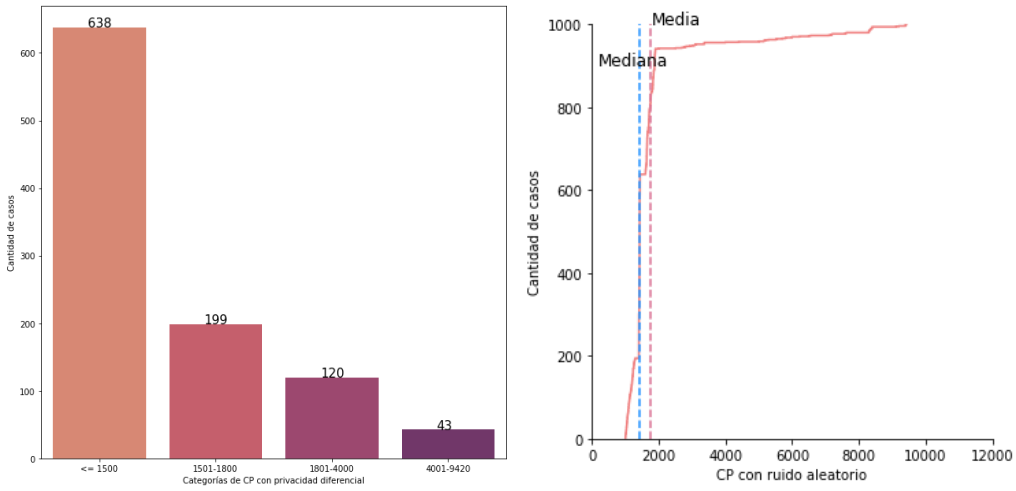
El algoritmo expresado en la figura 5 se programó con *Python*, parametrizándolo con un valor épsilon (ε) igual a 0.25. La elección de este valor fue en base a los resultados obtenidos en el apartado anterior donde el error fue menor del 5% para un tamaño de muestra de 950 resultando aceptable. Llevando a cabo la implementación de este, a continuación, se puede observar la evaluación de resultados de manera gráfica.

Gráfico 8: Distribución y distribución acumulada de individuos según rangos de código postal sin ruido aleatorio (variable original)



Fuente: elaboración propia con *Python*

Gráfico 9: Distribución y distribución acumulada de individuos según rangos de código postal con ruido aleatorio (variable distorsionada)



Fuente: elaboración propia con *Python*

De la comparación entre los gráficos 8 y 9 puede observarse que la distribución del código postal en categorías (gráficos de barras en cada caso) no se ve afectada por haber aplicado ruido aleatorio. De hecho, se mantiene la misma cantidad de individuos contenidos en cada

clase. Además, si se observa el comportamiento de la función de distribución (gráficos de la derecha en cada caso) puede notarse que son prácticamente iguales. Al observar numéricamente las estadísticas resumen (con líneas punteadas en los gráficos) en la tabla 3 a continuación, puede notarse como no se vieron afectados severamente sus valores.

Tabla 3: Medidas estadísticas resumen del código postal (CP) con y sin ruido aleatorio

Medidas	CP sin ruido aleatorio	CP con ruido aleatorio
Moda	1425	1425.77448
Media	1741.804	1742.578480
Mediana	1428	1428.774480

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

Para poder comprobar si existe homogeneidad de las varianzas en la distribución del código postal con y sin ruido aleatorio, se realiza una prueba de hipótesis de Levene[53]. Las hipótesis que se plantean son:

$$\begin{cases} H0: la\ varianza\ del\ código\ postal \text{ con y sin } ruido\ aletorio\ son\ iguales \\ H1: la\ varianza\ del\ código\ postal \text{ con y sin } ruido\ aletorio\ son\ distintas \end{cases}$$

Criterio de decisión para rechazar H0: p-valor < alfa=0.05

De llevar a cabo la prueba de Levene con *Python* (Ver código en Apéndice – sección "7. Implementación final de ruido aleatorio y evaluación de resultados" – sección 7.2) se obtiene un p-valor mayor a 0,99. Por tanto no se rechaza la hipótesis nula (H0). De esta manera es posible concluir que hay evidencia suficiente para considerar que las varianzas de la distribución del código postal con y sin ruido aleatorio son iguales. Ello implica que la distribución del código postal con ruido aleatorio no se vio significativamente modificada con un 95% de confianza, como también pudo observarse gráficamente.

Finalmente, si se toma el atributo código postal con ruido aleatorio y se intenta identificar al mismo individuo mostrado en el ejemplo del apartado 3.4.4, ya no es posible. Esto se debe a que como resultado de haber aplicado distorsión al código postal ahora toma el valor de 1002.77448 siendo distinto al valor original que era 1002. De esta manera no es posible realizar al reidentificación del individuo a partir de cruzar dos bases de datos donde una de

[53] Acerca de la prueba de hipótesis de Levene ver Bisquerra, R. (1987) "La prueba de Levene para la homogeneidad de varianzas en el BMDP" Revista de investigación educativa. 1987, v. 5, n. 9; p. 79-85.

ellas posee el atributo código postal distorsionado, ya sea individualmente o combinado con otros.

En función de los resultados obtenidos se logra conseguir datos distorsionados sin que se afecte significativamente la distribución del atributo código postal. De esta manera, es posible contar con información sobre este de forma protegida a la vez que utilizable en cualquier análisis posterior. La protección (o distorsión) alcanzada logra evitar la identificación de un individuo lo que brinda garantía sobre su privacidad en los términos definida en este trabajo. De este modo, es posible decir que en base a la aplicación realizada se logra mostrar la efectividad de aplicar ruido aleatorio por el mecanismo de Laplace propuesto por la metodología de Privacidad Diferencial para la protección de datos personales.

Conclusión del capítulo

En el presente capítulo, se ha propuesto a la metodología de la Privacidad Diferencial como aquella que permite brindar garantía de protección de datos personales en un contexto organizacional. A partir de la exposición de su concepto y fundamentos teóricos, se presentó como el método permite mitigar el riesgo de privacidad asociado a la utilización de datos. Mediante el desarrollo un caso de aplicación, se expuso su efectividad en la protección de un atributo personal contenido en una base de datos y como ello evita la identificación de un individuo perteneciente a la base utilizada.

La importancia de poder contar con una metodología de este tipo es que permite construir privacidad de datos personales desde el diseño de los sistemas de procesamiento de estos datos. Pero, también, permite exponer la sensibilidad de los datos personales como elemento clave para la construcción de la privacidad por defecto en un contexto organizacional. De esta forma, es posible abordar a la privacidad de datos personales en los términos en que fue definida en este trabajo. Se podrá llevar a cabo un accionar proactivo y preventivo para complementar a la legislación que se la considera de carácter reactivo.

Para poder llevar a cabo la presentación de la metodología de Privacidad Diferencial, en primer lugar, se expuso su definición conceptual. Por esta, se entiende al método que permite aplicar cierta distorsión a los datos personales pero que no invalida su utilidad posterior. De este modo, es posible que una organización pueda ejercer un uso responsable de datos personales al brindar garantías de protección a sus titulares. La posibilidad de contar e

implementar este tipo de metodología permite ejercer un control sobre el riesgo asociado al uso de este tipo de datos.

El control del riesgo de privacidad en el uso de datos personales en contextos organizacionales está asociado a la posibilidad de identificar o re identificar individuos en una base de datos. Como fue ejemplificado en segundo lugar, la metodología de la privacidad diferencial permite controlar el nivel de distorsión a implementarla a través de una función aleatoria sobre el verdadero valor de un dato. De esta manera, es posible lograr que no se exponga la información real de los individuos. Pero al mismo tiempo no invalida que los datos puedan ser procesados y analizados para la construcción de información. Tal posibilidad surge a partir de controlar un parámetro (épsilon). Este es el que permite evaluar cuanta distorsión es necesaria aplicar para evitar que un usuario del dato no logre identificar o re identificar individuos en su posterior uso. Así, es posible decir que la metodología presentada permite definir cuanto ruido aleatorio es necesario aplicar a los datos para alcanzar un nivel de protección adecuado o responsable como garantía de privacidad. En este sentido, su utilización resulta prometedora.

Si bien, este primer resultado alcanzado a partir de la ejemplificación de un caso predispone un escenario positivo para contribuir en la construcción de una gestión responsable dentro de un contexto organizacional, también muestra que por sí solo solo no será suficiente. En una última etapa de desarrollo del presente capítulo, a través del análisis de caso realizado, se muestra que los servicios *web* de consulta ofrecidos por las organizaciones estatales facilitan el acceso irrestricto a cierta información de los individuos a partir de contar con un identificador único. Frente a ello será necesario que se contemple aplicar medidas más robustas en la publicación de cierta información, que se definan principios de acción a seguir. En este sentido, la responsabilidad necesaria a asumir por las organizaciones, no solo se alcanzará a través de la aplicación de una metodología cuantitativa. Si no que también resultará de una revisión de las medidas adoptadas para la publicación de información sensible.

Para poder arribar a la conclusión anterior, en primer lugar, se debió construir manualmente una base de datos personales. Esto fue posible partiendo de un padrón de CUIT tomado de un repositorio online de datos abiertos del Gobierno de la Ciudad de Buenos Aires. A partir de este atributo único, se identificó individuos según género creando un nuevo atributo personal. Luego se tomó una muestra aleatoria de 1000 individuos. Tomando cada uno de

estos, se consultaron dos servicios *web* gratuitos que permitieron obtener varios atributos personales más, conformando así la base de datos trabajada.

A partir de la conformación de la base datos personales, en una primera instancia se anonimizó el CUIT (Clave Única de Identificación Tributaria). Sin lugar a duda, para evitar la exposición de identificadores únicos como este, es viable aplicar una técnica como la anonimización. Pero, aun así, de la combinación realizada de otros atributos contenidos en la base de datos se deriva que continúa existiendo el riesgo de reidentificación de individuos. Frente a esta situación la metodología de privacidad diferencial toma mayor relevancia. Su efectividad en términos de distorsionar un atributo personal permitió observar cómo se evita la identificación de un individuo. Mediante su aplicación para distorsionar un atributo (el código postal) que previamente se mostró que combinado con otros permitía identificar a un individuo, se realizó la evaluación de resultados. Con una conclusión favorable sobre su nivel de efectividad, se logra impedir la reidentificación de un individuo en la base de datos trabajada. Además, permite contar con el atributo distorsionado, pero sin que se pierda su utilidad para la construcción de información agregada y análisis posterior. De este modo se alcanza cierta garantía de privacidad para los individuos contenidos en la base de datos.

En base a la aplicación realizada queda al descubierto como la falta de una gobernanza completa de datos personales puede poner en riesgo la privacidad de los individuos. No solo surge como consecuencia de exponer identificadores únicos como el CUIT. Si no también de como la recombinación de atributos personales deriva en la identificación de los individuos. Si bien la utilización de información personal puede resultar necesaria para la realización de políticas públicas en pos de la construcción de bienestar social, no debería ser tratada solo en base a esta necesidad. El enfoque de evaluación de riesgos asociados implica que sea ampliado a considerar al ciudadano como parte del proceso.

De esta manera, el manejo de datos personales por parte de organizaciones públicas estatales invita a cuestionar cual es el grado necesario de responsabilidad que ello conlleva. A partir de ello se abre la posibilidad de plantear cuáles son los elementos estructurales que permitan establecer políticas y procesos concretos para llevar adelante el diseño de una gestión responsable de datos personales. Particularmente, el respeto al derecho de privacidad derivada del uso de datos personales deberá convertirse en esencial en un contexto social de fuerte digitalización. Esto permitirá encaminar hacia una cultura de responsabilidad organizacional entorno al uso de datos en el marco de un sistema social democrático. En el

siguiente capítulo se abordará la elaboración de una estructura de acción para guiar a las organizaciones hacia una gestión responsable de datos personales.

Capítulo 4: La responsabilidad como elemento principal en el tratamiento de la privacidad de datos personales en organizaciones

Introducción

La asimetría de poder producida en el mercado de datos como se argumentó en el capítulo 1, derivó en la necesidad de elaborar una definición de la privacidad de datos personales que incluya la responsabilidad organizacional, como se expuso en el capítulo 2. La insuficiencia de la regulación allanó el camino para argumentar sobre la necesidad de utilizar un enfoque desde el diseño y por defecto para abordar la problemática presentada. A su vez, esto implicó la necesidad de utilizar una metodología cuantitativa que brinde garantías de protección de datos personales. Como se expuso mediante un caso de aplicación en el capítulo 3, la Privacidad Diferencial resulta adecuada en este sentido.

Ahora bien, la actuación por defecto que argumenta sobre la necesidad de la responsabilidad organizacional no solo es una cuestión de decisión o teórica. También, expone la necesidad de poder cuantificarla para llevar a cabo una gestión operativa, es decir, basada en resultados. Poder medirla adecuadamente, abre la posibilidad de dimensionar su importancia (AEPD, 2019, 2020; Hoepman, 2022; Verhulst, 2022). Para abordar este desafío, se requerirá de operacionalizar los elementos que surgen como claves en el proceso de datos.

En este contexto, considerando el marco expuesto en capítulos anteriores, surge el siguiente interrogante: ¿en qué grado la responsabilidad organizacional contribuye en la construcción de privacidad? Para responder esta pregunta el objetivo a desarrollar en el presente capítulo consiste en elaborar un modelo que permita medir el grado de asociación entre privacidad y responsabilidad. Al partir de un marco de gestión de la privacidad sustentado en el enfoque desde diseño y por defecto, luego se definen las tareas operativas asociadas al cumplimiento de principios para dimensionar la responsabilidad desde una perspectiva operativa. De este modo se podrá identificar y variabilizar los elementos que surgen como claves para llevar a cabo una modelización.

A fin de elaborar un modelo que permita medir el grado de asociación entre responsabilidad y privacidad, el presente capítulo se divide en tres apartados. En el primero se desarrolla la conceptualización de la responsabilidad desde una perspectiva operativa. Para ello se elabora

el diseño de una estrategia de gestión responsable sustentada en los principios fundacionales de la privacidad como los presentados en el capítulo 2. Sobre esta base, y considerando el enfoque desde el diseño y por defecto, se especifican las tareas operativas asociadas al cumplimiento de un accionar responsable en base a datos. Finalmente se elabora un mapa que expone cuales actividades requieren un abordaje de definiciones conceptuales y cuales un enfoque cuantitativo para la aplicación de métodos de control.

La segunda sección, se centra la variabilización de los elementos que surgen como claves en el proceso llevado a cabo. Como resultado de la aplicación realizada en el capítulo 3, surgen tres elementos claves que son Épsilon (nivel de pérdida de privacidad), Error (nivel de distorsión generado en los datos) y el tamaño de la muestra. Luego, en un sentido operativo presentado en este capítulo, surge la Responsabilidad como el cuarto elemento clave. A partir de selección de estos elementos, se los operacionaliza a través de la construcción de variables, considerando a la Privacidad como una variable objetivo. Finalmente, se realiza un análisis estadístico para evaluar cual es el sentido de la relación entre las variables y si existe relación lineal entre ellas.

En la última sección, se desarrolla la propuesta de un modelo logístico para vincular a la privacidad con los elementos claves variabilizados. Se comienza por la presentación teórica del modelo, para derivar en la construcción del modelo específico. Finalmente se lleva a cabo su aplicación y se evalúan los resultados obtenidos en diferentes escenarios considerados. Como resultado se obtiene que la responsabilidad organizacional es una variable relevante en la construcción de privacidad de datos personales. Si bien, no se abordan todas las posibilidades existentes, es un primer resultado valioso ya que permite interpretar la importancia de su contribución. Además, se constituye en una herramienta de gestión que puede ser aplicada en organizaciones públicas para llevar a cabo una estrategia de gestión responsable de la privacidad.

4.1 La gestión de la responsabilidad para alcanzar privacidad

En la conformación de un mercado de datos como fue presentado en el capítulo 1, las organizaciones tienen la particularidad de basar sus decisiones en evidencia, en información. Para que este proceso resulte acorde requiere de ser gestionado. Por un lado, para poder lograr la integridad de los datos necesarios de forma tal de alcanzar calidad y confiabilidad para la obtención de resultados. Por el otro, dado el poder generado que limita la capacidad de acción del individuo y la normativa existente como fue presentado en el capítulo 2, surge

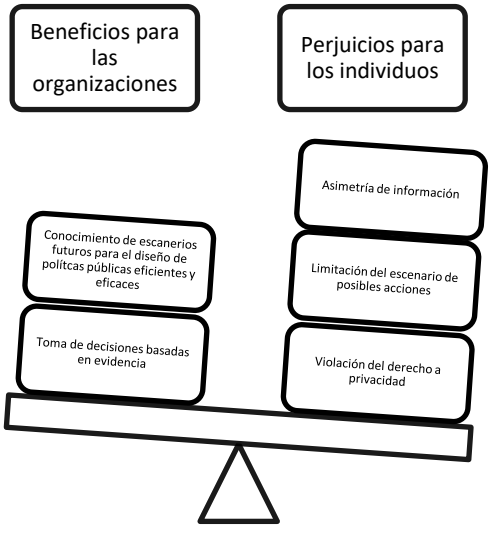
que la responsabilidad organizacional es un elemento central que también requiere de ser gestionado.

La gestión de la responsabilidad puede ser abordada desde un accionar operativo. Esto permite pasar de la teoría a acciones concretas (AEPD, 2019, 2020; Hoepman, 2022; Verhulst, 2022). Por esta razón, la cuantificación de elementos que resultan claves, cobra vital importancia para que puedan ser gobernados adecuadamente. En particular, en el caso de las organizaciones públicas se torna fundamental para contribuir al bienestar social.

4.1.1 La gestión de la responsabilidad desde un enfoque operativo en contextos organizacionales

En línea con el enfoque propuesto a lo largo del capítulo 1 de este trabajo, Verhulst (2022) en su trabajo "*Operationalizing digital self-determination*", sostiene que, para lograr mitigar la asimetría de información producida en el mercado de datos, limitar el poder de acción organizacional y evitar la violación de la privacidad, es fundamental que "la teoría se traduzca en implementación práctica" (pag.9). A este fin, adoptar principios que aporten los conceptos principales luego permitirá definir los pasos a seguir. Esto resultará de igual importancia que poder comprender como llevarlo a la práctica de manera responsable. En definitiva, se trata de ir en búsqueda de un equilibrio entre las potencialidades que el uso de los datos personales brinda a las organizaciones y los perjuicios que podría ocasionarle a los individuos.

Figura 6: Beneficios y perjuicios en el uso de datos personales en la sociedad de la información

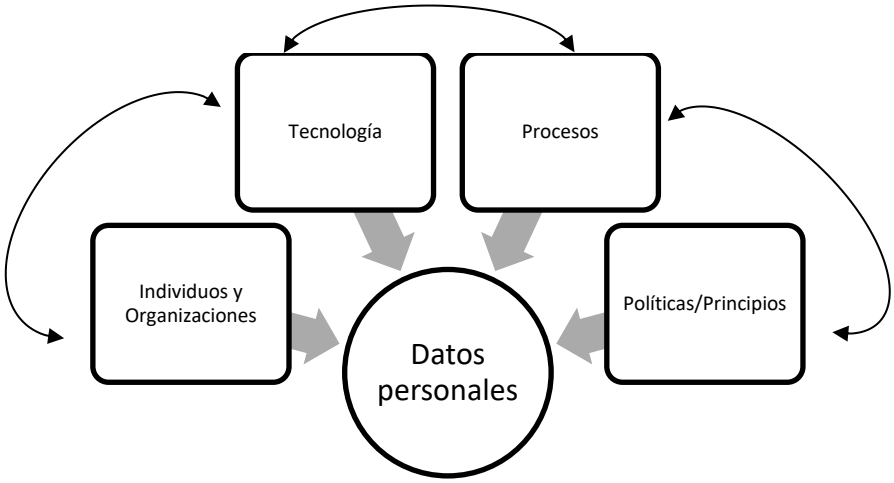


Fuente: elaboración propia

Ante la posibilidad de un desequilibrio entre beneficios y perjuicios derivados del uso de datos personales, resultará necesario llevar adelante el diseño de una estrategia de gestión responsable de la privacidad de datos en un contexto organizacional. Por un lado, permitirá que se cumpla con cierta normativa existente. Pero como fue planteado en el capítulo 2 del presente trabajo, en términos de prevención de perjuicios derivados, el solo cumplimiento de una normativa puede resultar insuficiente. Por esta razón, resulta valioso adoptar una perspectiva desde el diseño y por defecto para abordar a la privacidad de datos. De este modo, la responsabilidad organizacional será incorporada por defecto desde el momento en que se diseñan los procesos de datos.

El escenario que plantea el enfoque anterior – basado en lo desarrollado en los capítulos 1 y 2- implica partir de considerar cuatro partes clave asociadas (Cavoukian, 2011; AEPD, 2019, 2020; Hoepman, 2022; Verhulst, 2022) al uso de datos para comenzar a diseñar una estrategia de gestión de privacidad:

Figura 7: Aspectos para el diseño de una estrategia de gestión de privacidad



Fuente: elaboración propia

Los aspectos considerados en la figura 7, surgen de la relación de interacción entre individuos generadores de datos y las organizaciones en un espacio digital. Este es creado por estas últimas y se denomina plataforma[54]. En este entorno, las organizaciones utilizan herramientas tecnológicas para recopilar los datos personales y realizar su posterior procesamiento para la construcción de información. A partir de ello, surge la creación de un proceso continuo que requerirá de una gestión responsable basada en políticas y principios

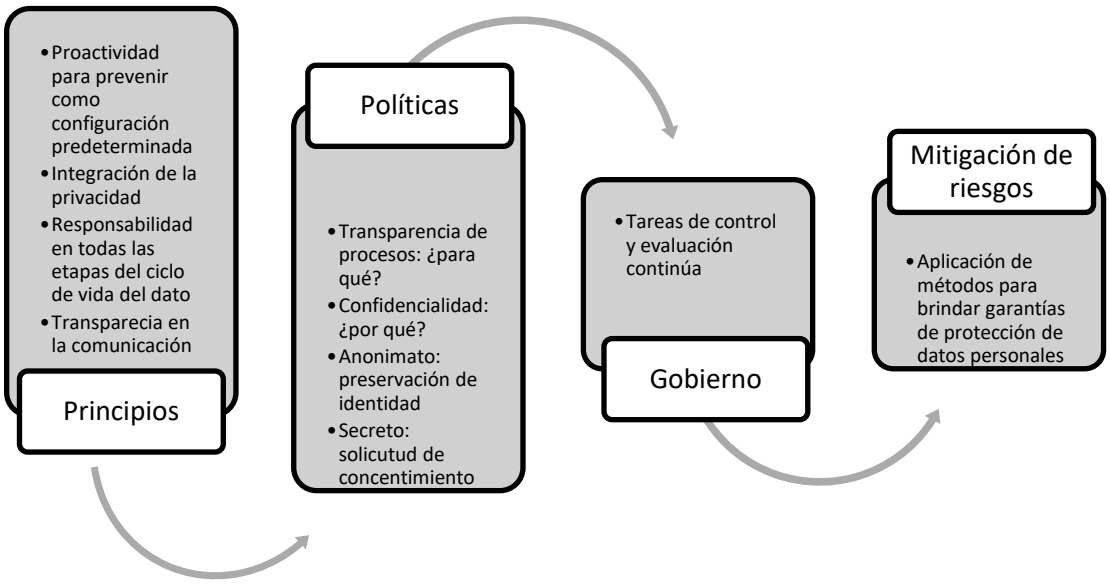
[54] Definidas en el apartado 1.2 como el espacio digital de interacción entre individuos y organizaciones, que en el caso de del Estado recibe el nombre de gobierno digital.

claros para mitigar los riesgos asociados a privacidad. En este sentido, los principios fundacionales presentados en el capítulo 2, pueden considerarse como eje central para adoptar una definición de privacidad de datos personales que contemple a la responsabilidad organizacional.

Una definición de privacidad de datos personales en línea con lo mencionado en el párrafo anterior es la propuesta en este trabajo. En el apartado 2.1.1 se la definió como la posibilidad de que un individuo generador de datos al interactuar con una organización en un contexto de red digital pueda obtener garantías suficientes de que sus datos están siendo debidamente protegidos dentro de un marco de responsabilidad. Desde esta perspectiva, la responsabilidad surge como un elemento clave que deberá ser considerado y gestionado dentro de la organización.

Ahora bien, para pasar de la definición a la aplicación práctica construyendo un vínculo entre procesos y principios aplicados a las herramientas tecnológicas utilizadas, deberán definirse acciones operativas o tareas específicas que contribuyan a lograrlo. De este modo, se entrará en la fase de diseño operativo de gestión de la privacidad de datos bajo un paragua de responsabilidad. Tales tareas deberán facilitar encontrar un equilibrio entre la cantidad de datos utilizadas y el anonimato, el secreto, la transparencia y la confidencialidad definidos en el apartado 2.3. De esta forma se podrán establecer los controles necesarios para mitigar el riesgo de identificación de individuos cuando los datos personales son utilizados.

Figura 8: Plan estratégico de gestión de privacidad



Fuente: elaboración propia

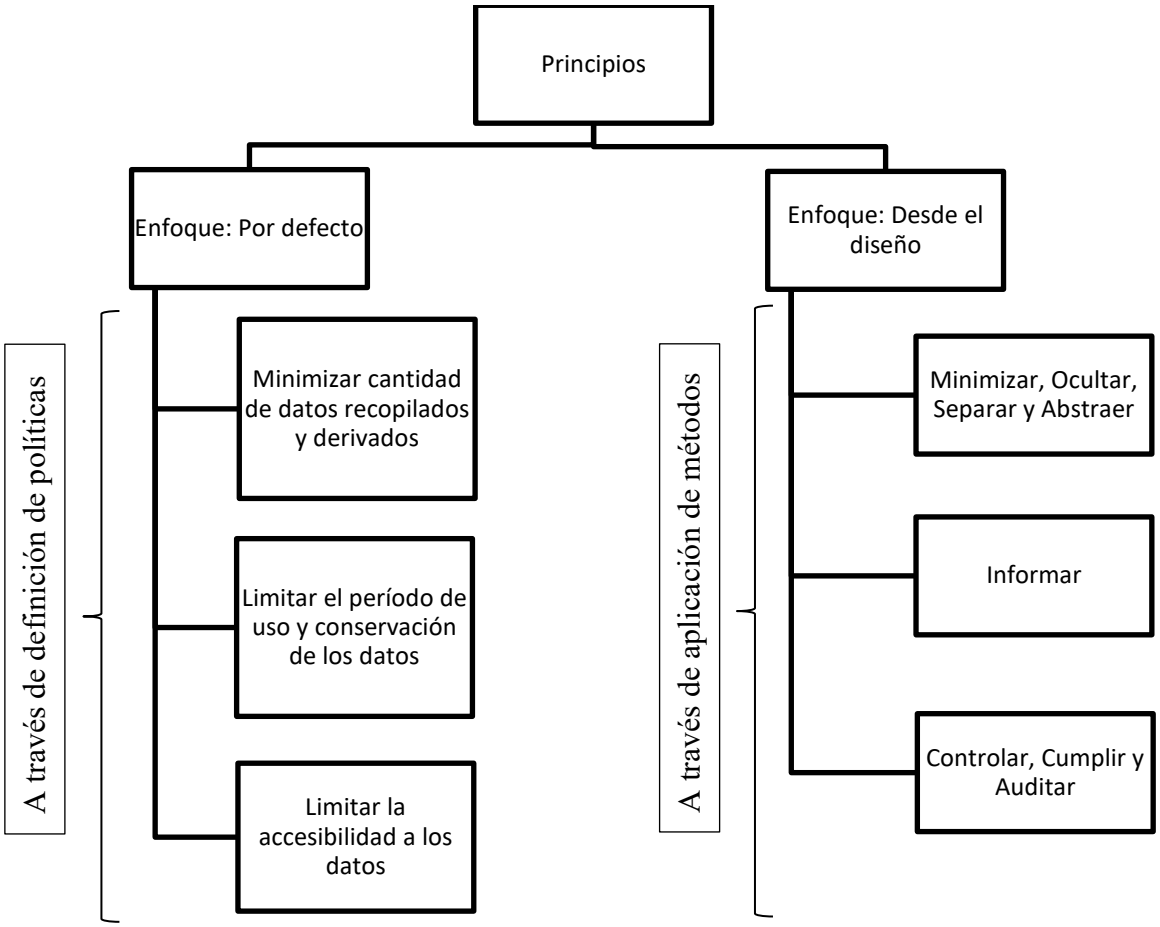
En la elaboración de una estrategia de gestión como la presentada en la figura 8, si bien los principios son enunciaciones teóricas, permitirán establecer los lineamientos a seguir a la hora de elaborar las políticas para gobernar responsablemente los procesos aplicados a través de tecnología. Esto ayudará a llevar adelante todas las tareas necesarias en cada etapa del ciclo de vida del dato en la organización con el fin de lograr mitigar los riesgos asociados a privacidad de datos. Así se podrá lograr brindar garantías de protección de datos a sus titulares. De este modo, el diseño de un plan estratégico de gestión de privacidad de datos basado en estos cuatro aspectos principales encontrará a la responsabilidad organizacional como punto de partida constituyéndose en un factor clave.

Ahora bien, para poder aproximarse a dimensionar la contribución operativa de la responsabilidad en la construcción de privacidad de datos, resulta necesario especificar cuáles son las principales acciones de gestión a realizar. Estas deberán ser aquellas que operativamente permitan contribuir al cumplimiento de los principios establecidos a través de la obtención de resultados concretos. La elaboración de todas ellas en conjunto permitirá posteriormente construir una variable que exprese si hay cumplimiento o no.

4.1.2 Principales acciones operativas

Para poder alcanzar el cumplimiento de los principios fundacionales de la privacidad definidos en el apartado 2.3.2, las acciones a realizar se pueden agrupar en dos grupos (AEPD, 2019, 2020). Un primer grupo asociado a la privacidad por defecto y un segundo grupo asociado a la privacidad desde el diseño. En el primero se encontrarán aquellas asociadas a minimizar la cantidad de datos recopilados y brindar transparencia al titular de los datos. En el segundo, se encontrarán las acciones operativas destinadas a mitigar el riesgo de identificación de individuos en base al uso de datos personales. En este último, se requerirá de la selección de una metodología cuantitativa para brindar garantías de protección de datos.

Figura 9: Acciones para el cumplimiento de los principios fundacionales de privacidad



Fuente: elaboración propia

En la figura 9, se puede apreciar dos grupos de acciones necesarias. En el caso de las acciones por defecto se trata de definiciones a través de la elaboración de políticas para guiar su cumplimiento dentro del contexto organizacional. Por otro lado, las acciones involucradas en el diseño, requerirá de la adopción de métodos cuantitativos para poder cumplimentar con cada una de ellas.

La minimización de datos recopilados refiere a que cuando se le solicitan datos al individuo, estos sean los mínimos indispensables al fin determinado. Pero, como también es posible construir nuevos datos en base a los originales, será relevante tenerlo en cuenta. Por ejemplo, a partir de la fecha de nacimiento es posible determinar la edad de una persona, entre otros. En este caso, es recomendable adoptar un criterio de generalización. En lugar de utilizar la edad exacta, emplear rangos de edad. El criterio sobre la cantidad de datos deberá definirse minuciosamente para darle posibilidad al titular de estos de decidir si está dispuesto a

brindarlos. Además, para poder establecer qué nivel de distorsión deberá aplicarse a los datos personales para lograr protección como garantía de privacidad.

A su vez, deberá tenerse en cuenta establecer limitaciones de uso de los datos. Dentro de un contexto organizacional, existen diferentes áreas de trabajo. Cada una de estas, requiere el acceso a ciertos datos y no a todos. En este sentido, las configuraciones de seguridad interna deberán limitar el acceso de los usuarios creando perfiles según necesidades específicas. Además, será necesario limitar el tiempo de conservación de los datos en cada etapa de uso. Particularmente, la Ley 25.326 establece un período de conservación de los datos personales de como máximo 3 meses, a partir de que el destino de uso haya finalizado.

Entre las organizaciones públicas del estado es frecuente que se comparta el acceso a datos. En ocasiones con el fin de validar cierta información brindada por el individuo[55]. Aun así, se requiere limitar el acceso para no incurrir en abuso que conlleve a una violación de la privacidad[56]. Además, teniendo en cuenta que existe la Ley 27.275[57] de acceso a la información pública, resultará necesario establecer criterios para la entrega de este tipo de información. Para ello deberá considerarse que por más que ciertos datos pueden considerarse públicos, pueden vulnerar el derecho de privacidad de un individuo. Por ejemplo, cuando se entrega información transaccional de viajes en transporte público -a excepción de los casos solicitados vía orden judicial-, si la misma no está debidamente acotada, agregada, podría identificarse los recorridos que realiza cada individuo y en consecuencia localizarlo. También, los identificadores directos como el CUIT ponen en riesgo la privacidad de un individuo, como fue mostrado en el desarrollo llevado a cabo en el capítulo 3 del presente trabajo. Sin duda, deberá establecerse minuciosamente que información podrá ser pública (accesible) y cual no.

Este primer grupo de acciones permitirá guiar hacia la construcción de una gestión responsable por parte de las organizaciones. A su vez, como fue deslizándose en lo mencionado anteriormente, comienza a surgir la necesidad de llevar a cabo tareas específicas para alcanzar resultados concretos. En esta línea, surgen aquellas que se agrupan en el segundo grupo bajo el enfoque desde el diseño. Particularmente, se trata de alcanzar

[55] Por ejemplo, validar el número de DNI en RENAPER o CUIT en AFIP.
[56] Como por ejemplo el caso mencionado en el apartado 2.2.3 sobre Ministerio de Justicia y Seguridad de la Ciudad Autónoma de Buenos Aires
[57] En línea: https://servicios.infoleg.gob.ar/infolegInternet/anexos/265000-269999/265949/norma.htm

objetivos de privacidad relacionados con desvinculación, control y transparencia de datos (APDE, 2019; Hoepman, 2022).

Por desvinculación se hace referencia a que a partir del uso de los datos personales se impida la identificación de su titular. Mediante la acción de minimizar, como fue mencionado anteriormente, se busca que la cantidad de datos utilizados sea la menor posible de modo tal de disminuir el riesgo de identificación. En este sentido, la selección adecuada, la exclusión detallada y la eliminación post uso, resultarán acciones necesarias para alcanzar este fin. En cambio, el ocultamiento se vincula a acciones de disociación y ofuscamiento. En este punto, toma relevancia la selección adecuada del método cuantitativo a aplicar para lograr distorsionar los identificadores directos e indirectos. Particularmente en el capítulo 3 de este trabajo, se mostró como la metodología de Privacidad Diferencial permite cumplimentar con dicho fin mediante un caso de aplicación. Por esta razón, en principio, es posible recomendar su utilización. Además, esta metodología permite obtener dos elementos que resultan claves para llevar un control de la privacidad de datos: el nivel de distorsión a aplicar a través de su parámetro épsilon y la magnitud del error cometido.

La acción de separación de datos refiere a cómo deben ser almacenados de forma tal que se evite alcanzar un perfilamiento completo de un individuo (APED, 2019). En este sentido, se suele recomendar que se utilicen diferentes bases de datos para almacenamiento interno y que a su vez se apliquen procesos independientes. Por otra parte, la acción de abstracción refiere a la utilización de los datos de manera agregada o agrupada, como se mencionó anteriormente. También, de manera distorsionada, es decir, aplicando cierto ruido aleatorio que evite el uso del dato real pero que no impida su validez para un análisis posterior.

El control, el cumplimiento y la auditoría, son acciones que refieren al objetivo de control de los procesos que se aplicarán a los datos, pero que a su vez están directamente ligados al objetivo de informar (APED, 2019). Lo que se busca mediante estas acciones es que la organización lleve adelante una comunicación transparente sobre la utilización de los datos personales. Para ello, explicitar de manera clara y entendible cuáles son los métodos de control que se emplean, así como se lleva a cabo el cumplimiento de la normativa existente y el proceso de rendición de cuentas, contribuye a que los individuos puedan ganar confianza para brindar sus datos. Para lograrlo, llevar adelante una documentación detallada de lo realizado bajo estas acciones suele ser necesario y la forma frecuente de realizarlo.

De este modo, la organización podrá transparentar su accionar al poder brindar información de manera responsable sobre el cumplimiento de todas las acciones mencionadas. Al mismo tiempo permitirá reducir la asimetría de información existente a través de elaborar un documento y ponerlo a disposición públicamente. Esto, normalmente, se conoce como políticas de privacidad y deberá ser de fácil lectura y comprensión para cualquier individuo.

Una vez que se han descripto las acciones que permiten construir un accionar responsable en un contexto organizacional, resulta importante poder dimensionar su contribución a la construcción de privacidad de datos. Esto es, poder modelizar el vínculo entre cumplimiento del conjunto de todas las acciones y la privacidad mencionada. Para ello, en primer lugar, deberá poder determinarse cuáles son los elementos claves que intervienen operativamente.

4.2 Los elementos claves para el diseño de una estrategia responsable de privacidad

En este apartado, se presentan a los elementos claves que permiten evaluar el grado de contribución en la construcción responsable de privacidad de datos. Constituyen características esenciales dentro del proceder operativo propuesto en lo desarrollado en este trabajo a partir del enfoque de privacidad desde el diseño y por defecto. Un tipo de elementos está asociado a los parámetros relevantes de la metodología cuantitativa seleccionada para distorsionar datos personales. Otro tipo está asociado al cumplimiento de las acciones operativas de gestión responsable descriptas en el apartado anterior. Ambos tipos se pueden conformar en variables que a su vez se constituyen como las dimensiones claves para evaluar el tipo de relación con la privacidad de datos.

4.2.1 Selección de elementos claves

Una vez que se cuenta con una estrategia de diseño de privacidad que integra a la privacidad como parte de los procesos de datos como la propuesta en el apartado 4.1, se requerirá que los datos personales sean protegidos. Por un lado, en el caso de Argentina, para cumplimentar con legislación vigente. Por el otro, para construir un accionar responsable por parte de la organización. Esto expone, en primer lugar, la necesidad de seleccionar un método cuantitativo para la protección de datos personales. Para ello, y como fue propuesto en el capítulo 3 del presente trabajo, se propone la utilización de la metodología de la Privacidad Diferencial.

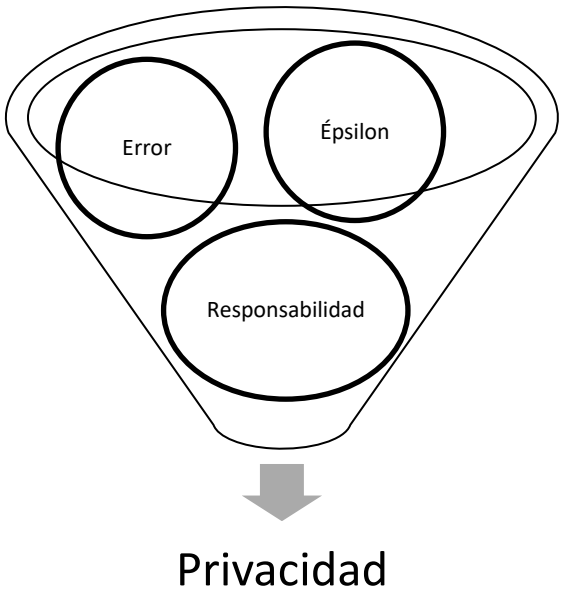
A partir de los resultados obtenidos en el apartado 3.4.5, fue posible obtener datos distorsionados sin que se afecte la distribución de un atributo personal en comparación con

sus valores originales. Así la utilización de datos personales es posible y bajo un accionar responsable ya que brinda garantía de protección a sus titulares en la medida que como resultado se imposibilita la identificación del individuo. Al mismo tiempo, para quien lleve adelante la aplicación de este, le permitirá controlar el nivel de privacidad logrado a través de dos parámetros claves: épsilon y el error cometido. Dada la dimensión de su importancia, ambos resultan esenciales para la construcción de privacidad de datos. En el caso de épsilon porque mide en nivel de pérdida de privacidad; en el caso del error porque mide cual es, en términos proporcionales, la distancia entre el verdadero valor y el valor distorsionado. Por esta razón se los selecciona como elementos claves a variabilizar.

El otro elemento clave es la responsabilidad, como fue expuesto en el apartado 4.1 del presente capítulo. A partir de considerar el cumplimiento de cada acción definida, puede agruparse en un único elemento clave que es la responsabilidad. Al ser operacionalizada de esta manera, permitirá que dentro del contexto organizacional un responsable pueda llevar adelante un proceso de evaluación de cumplimiento. Al mismo tiempo, permite convertirla en auditable pero también en transparente. No solamente por cumplimentar con la normativa vigente sino para comunicarlo de manera simple a los individuos.

Por todo lo expuesto hasta el momento, es posible determinar que los elementos que resultan claves en la construcción de la privacidad para una gestión responsable de datos personales son:

Figura 10: elementos claves para la construcción de privacidad



Fuente: elaboración propia

Este relevamiento reflexivo y deductivo puede operacionalizarse a través de variables cuantitativas. La importancia de ello radica en poder analizar qué tipo de relación existe y en que cuantía contribuyen para la construcción de la privacidad de datos.

4.2.2 Operacionalización de los elementos claves: creación de variables

Tomando los elementos claves seleccionados en el apartado anterior, en este apartado se especifica como se transforman en variables a ser analizadas. Esto permitirá construir un conjunto de datos conformado por cuatro variables: épsilon, error, responsabilidad y privacidad de datos. A partir de ello será posible analizar cuantitativamente la relación existente entre ellas.

Para lograr conformar una variable con valores de épsilon y obtener valores del error cometido al aplicar la metodología de Privacidad Diferencial, se aplica el algoritmo definido en el apartado 3.4.5. Para ello se define una serie de valores aleatorios de épsilon conformada por: [0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.05]; y una serie de tamaños de muestra [100, 150, 200, 250,300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950]. Utilizando el mismo conjunto de datos que en el mencionado apartado y tomando la misma variable código postal, se realiza una corrida del algoritmo (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "2. Modelización de la privacidad y responsabilidad"). Como resultado se obtiene por cada muestra y valor de épsilon, el error cometido en términos porcentuales. De este modo se cuenta con 342 valores posibles. A continuación, una muestra de los datos:

Tabla 4: variabilización de épsilon y error

Tamaño de muestra	Épsilon	%Error
100	0.95	4.0
100	0.9	4.0
100	0.85	5.0
….	….	….
950	0.15	4.3
950	0.10	4.9
950	0.05	9.7

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

El siguiente paso, consiste en la construcción de la variable Responsabilidad. Esta podrá tomar cinco valores: 1, si se cumplió con la realización de todas las acciones descriptas en

el apartado 4.1.2; 0.75 si se cumplió en un 75%, 0.5 si se cumplió en un 50%, 0.25 si se cumplió en un 25% o 0 en caso de incumplimiento. Para abordar la opción de combinación de los valores anteriormente obtenidos de tamaño de muestra, épsilon y error y contar con todos los valores definidos de Responsabilidad, se duplican los resultados obtenidos en la tabla 4 y se crea la variable. En la tabla 5 a continuación se visualiza una muestra de los datos obtenidos:

Tabla 5: variabilización de épsilon, error y responsabilidad

Tamaño de muestra	**Épsilon**	**%Error**	**Responsabilidad**
100	0.95	4.0	0
100	0.95	4.0	0.25
100	0.95	4.0	0.5
100	0.95	4.0	0.75
100	0.95	4.0	1
…	…	…	…

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

Finalmente, se crea la variable Privacidad siendo la variable objetivo. Se define que podrá tomar valor 1 si se logra construir privacidad; en caso contrario valdrá 0. Para su creación, se consideran diferentes reglas en base a los valores de épsilon, error y responsabilidad obtenidos previamente. Esto es, se plantean diferentes escenarios con cierta simplificación a los fines de poder representar la mayor cantidad de posibles situaciones que pueden surgir. De este modo, se conforman tres conjuntos de datos, uno por cada regla aplicada para definir los valores de la variable privacidad. A continuación, las reglas para definir el valor de la variable privacidad son:

Tabla 6: reglas para definir el valor que toma la variable privacidad

Escenario	**Épsilon**	**%Error**	**Responsabilidad**	**Privacidad**
1	≤ 0.5	≤ 5	>=0.75	1; en otro caso 0
2	≤ 0.25	≤ 5	>=0.75	1; en otro caso 0
3	≤ 0.15	≤ 5	>=0.75	1; en otro caso 0

Fuente: elaboración propia

En cada escenario, el conjunto de datos conformado finalmente posee 4 variables y 1710 valores. Esto es, por cada combinación de tamaño de muestra, valor de épsilon, error obtenido y valor de responsabilidad, la privacidad vale 0 o 1. Valdrá 1 cuando se cumpla la

regla definida en la tabla 6. En otro caso, valdrá 0. A continuación se exponen muestras de algunos casos para visualizar que sucede con la variable Privacidad según las reglas establecidas en cada escenario:

Tabla 7: ejemplos de valores por escenario

Tamaño de muestra	Épsilon	Error	Responsabilidad	Privacidad
Escenario 1				
950	0.2	5.0	0	**0**
950	0.2	5.0	0.25	**0**
950	0.2	5.0	0.5	**0**
950	0.2	5.0	0.75	**1**
950	0.2	5.0	1	**1**
Escenario 2				
950	0.2	5.0	0	**0**
950	0.2	5.0	0.25	**0**
950	0.2	5.0	0.5	**0**
950	0.2	5.0	0.75	**1**
950	0.2	5.0	1	**1**
Escenario 3				
950	0.15	4.3	0	**0**
950	0.15	4.3	0.25	**0**
950	0.15	4.3	0.5	**0**
950	0.15	4.3	0.75	**1**
950	0.15	4.3	1	**1**

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

Para poder determinar si existe relación entre las variables explicativas creadas (tamaño de muestra, épsilon, error y responsabilidad) y la obtención de privacidad, en primer lugar, se realiza un análisis estadístico. En base a los resultados preliminares que se encuentren, luego se podrá definir si es posible la modelización de la relación. En caso de ser posible, se podrá proponer un modelo acorde a partir del cual se pueda evaluar operativamente la contribución de cada variable, y en particular de la responsabilidad, para la obtención de privacidad de datos.

4.2.3 Análisis estadístico inicial

Tomando como datos los obtenidos en el apartado anterior, se busca analizar en primer lugar, la relación de cada variable predictora con la variable objetivo. Para ello, se calcula la covarianza. La covarianza es una medida de asociación entre los valores de dos variables

(Canavos y Medal, 1987) que permite conocer cuál es el sentido (positivo o negativo) de la variación de los valores entre dos variables en caso de que exista relación entre ellas. Se define como:

$$\boldsymbol{Cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)]} \quad (10)$$

donde E es valor esperado, $\boldsymbol{\mu_x}$ media de X y $\boldsymbol{\mu_y}$ media de Y.

Al realizar el cálculo con *Python* (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "2. Modelización de la privacidad y responsabilidad"), se obtiene que:

Tabla 8: valores de la covarianza

	Escenario 1	Escenario 2	Escenario 3
Variables predictoras	Privacidad	Privacidad	Privacidad
Tamaño de muestra	13.136	7.139	2.838
Épsilon	- 0.02	- 0.017	- 0.005
Error	- 0.299	- 0.058	- 0.014
Responsabilidad	0.047	0.014	0.005

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

De la tabla 8, se puede observar que, en todos los escenarios planteados, la variable épsilon y error tienen una relación negativa (valor de covarianza negativo) con la variable privacidad; mientras que el tamaño de muestra y la responsabilidad tienen una relación positiva. Este resultado es esperable en la medida que a menor pérdida de privacidad (menor valor de épsilon y por tanto mayor distorsión aplicada en los datos) y mayor nivel de error (mayor distancia porcentual entre el verdadero valor de la variable y el valor distorsionado) implica una afectación negativa en la construcción de privacidad (mayor exigencia). En cambio, en el caso del tamaño de muestra (a mayor tamaño de muestra, menor error como se mostró en el apartado 3.4.5) y de la responsabilidad (cumplimiento de acciones), la relación muestra una contribución directa positiva a la obtención de privacidad de datos.

Ahora bien, con el fin de poder modelizar el vínculo entre las variables predictoras y la privacidad de manera lineal de forma tal de poder brindar una explicación lo más sencilla posible para contribuir a la transparencia, surge la pregunta de si efectivamente existe una asociación de este tipo entre las variables. Para poder medir el grado de asociación lineal entre las variables se utiliza el coeficiente de correlación de Pearson (Canavos y Medal, 1987). Este viene dado por:

$$\boldsymbol{\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}} \qquad (11)$$

donde $\boldsymbol{Cov(X,Y)}$ es la covarianza definida previamente, $\boldsymbol{\sigma_x}$ desvío de X y $\boldsymbol{\sigma_y}$ desvío de Y.

El coeficiente de correlación de Pearson es un valor que se encuentra entre -1 y 1. Su valor de -1 indica una relación lineal perfectamente negativa y su valor de 1 indica una relación lineal perfectamente positiva entre las variables. En caso de ser 0, indica ausencia de relación lineal. Llevando a cabo el cálculo con *Python* (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "2. Modelización de la privacidad y responsabilidad") se obtienen los siguientes resultados:

Tabla 9: valores del coeficiente de correlación de Pearson

	Escenario 1	Escenario 2	Escenario 3
Variables predictoras	Privacidad	Privacidad	Privacidad
Tamaño de muestra	0.153	0.1449	0.097
Épsilon	- 0.2259	- 0.2239	- 0.1573
Error	- 0.1043	- 0.0353	- 0.0143
Responsabilidad	**0.4012**	**0.2091**	**0.1211**

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

A partir de los resultados mostrados en la tabla 9, se destaca la alta correlación positiva entre la variable responsabilidad y privacidad. En el caso del escenario 1, es del 40% aproximadamente resultando muy superior al resto de las variables. Luego, cuando establece mayor exigencia en los umbrales para alcanzar distorsión de datos, si bien el valor de la correlación es de cuantía menor sigue siendo alto. Esto implica que existe una contribución lineal directa de magnitud importante del cumplimiento de la responsabilidad para la obtención de la privacidad de datos. Para poder determinar la significatividad estadística de este resultado se realiza una prueba de hipótesis. De llevar a cabo esta prueba con *Python* (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "2. Modelización de la privacidad y responsabilidad"), en todos los casos se rechaza la hipótesis de que la correlación no es significativa al obtener un p-valor cercano a cero. De esta manera, se considera que existe sustento suficiente para proponer una modelización del vínculo entre responsabilidad y obtención de privacidad a través de un modelo lineal.

4.3 Vinculación entre la responsabilidad organizacional y la privacidad de datos

En este apartado, se busca determinar en qué grado contribuye el cumplimiento de la responsabilidad en la obtención de privacidad. Para ello, se propone la utilización de un

modelo acorde a una variable binaria como es el caso de la Privacidad definida en el apartado anterior que permite vincular linealmente a las predictoras con la variable objetivo. De este modo se podrá poner a prueba la hipótesis de que la responsabilidad es un elemento clave para la obtención de privacidad de datos. A su vez, constituye una herramienta de gestión que podrá ser considerada para su aplicación en un contexto organizacional.

4.3.1 Modelización del vínculo entre responsabilidad y privacidad

En el capítulo 4.2, la privacidad fue definida como una variable discreta binaria que solo puede tomar valor 0 o 1. Este comportamiento particular, requiere de un modelo que sea capaz de transformar a contínua dicha variable para poder vincularla con las variables explicativas de manera lineal. Un modelo que permite realizar dicha transformación es el modelo de regresión logística.

La regresión logística tiene por objetivo describir la relación lineal entre una variable objetivo y un conjunto de variables independientes, como todos los modelos pertenecientes a la familia de los lineales en términos estadísticos (Agresti, 2012; Hosmer et al., 2013). A diferencia de un modelo de regresión lineal, en el modelo de regresión logística la variable objetivo es binaria. Por esta razón, adopta una forma diferente, aunque la estructura de análisis y evaluación es muy similar, lo que lo hace fácilmente interpretable.

La ecuación (Agresti, 2012; Hosmer et al., 2013) propuesta por el modelo de regresión logística viene dada por:

$$\ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \qquad (12)$$

donde β_i son los parámetros del modelo a partir de los cuales se relacionan las variables predictoras con la variable objetivo; x_i son las variables predictoras; $\ln\left(\frac{\pi(X)}{1-\pi(X)}\right)$ es la transformación *logit* mediante la cual se logra la linealización de la relación entre las variable objetivo (Y) y las predictoras; $\pi(X) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}$ es la probabilidad de ocurrencia de la Y condicionada a los valores de las x_i. A partir de aquí, el modelo estima el valor de la variable Y según la siguiente decisión: vale 1 si P(Y=1/x_i) >0.5, sino vale 0, donde 0.5 es el límite de decisión.

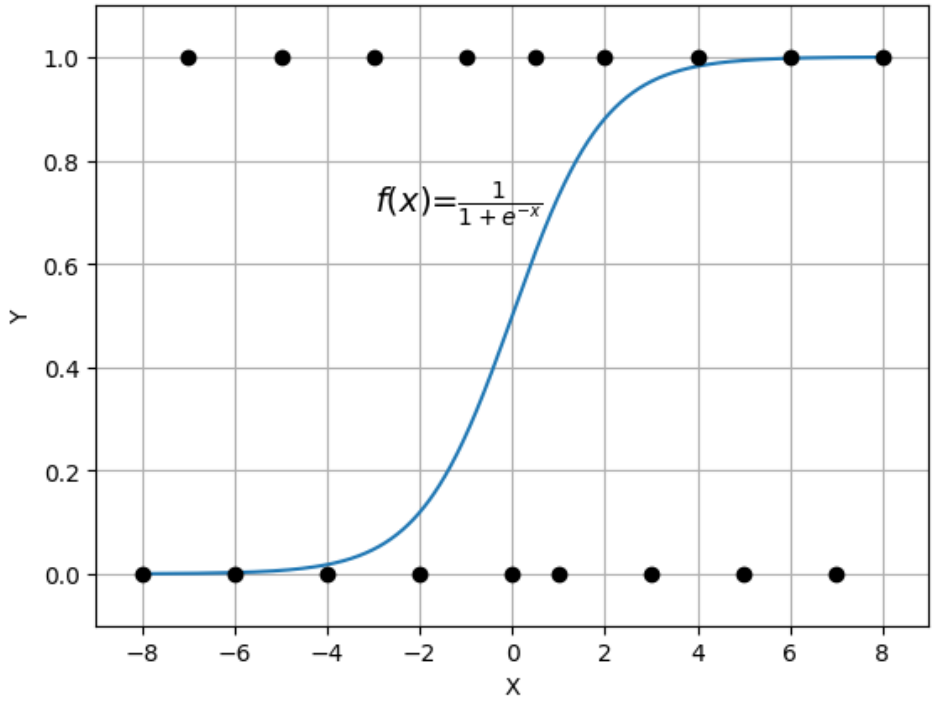
El *logit* surge de una transformación logarítmica aplicada al caso cuando la variable Y=1. Dado que, Y solo puede tomar valor 1 o 0, la proporción (*odd ratio*) de ocurrencia de Y=1 dado X viene dada por:

$$P(Y = 1/X) = \frac{\pi(X)}{1-\pi(X)} \quad (13)$$

donde $\pi(X)$ es la probabilidad de ocurrencia de Y=1 y 1-$\pi(X)$ es la probabilidad de ocurrencia de Y=0.

Al aplicar logaritmo natural a la ecuación 13, se transforma en contínua la distribución de la proporción obtenida que recibe el nombre *logit* (De Jong y Heller, 2008; Agresti, 2012; Hosmer et al., 2013). Este, resulta ser la inversa de la función matemática sigmoide (Jurafsky y Martin, 2024). La función sigmoide (f(x)=$\frac{1}{1+e^{-x}}$), gráficamente se representa como (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "1. Representación del proceso logístico"):

Gráfico 10: representación de la función sigmoide

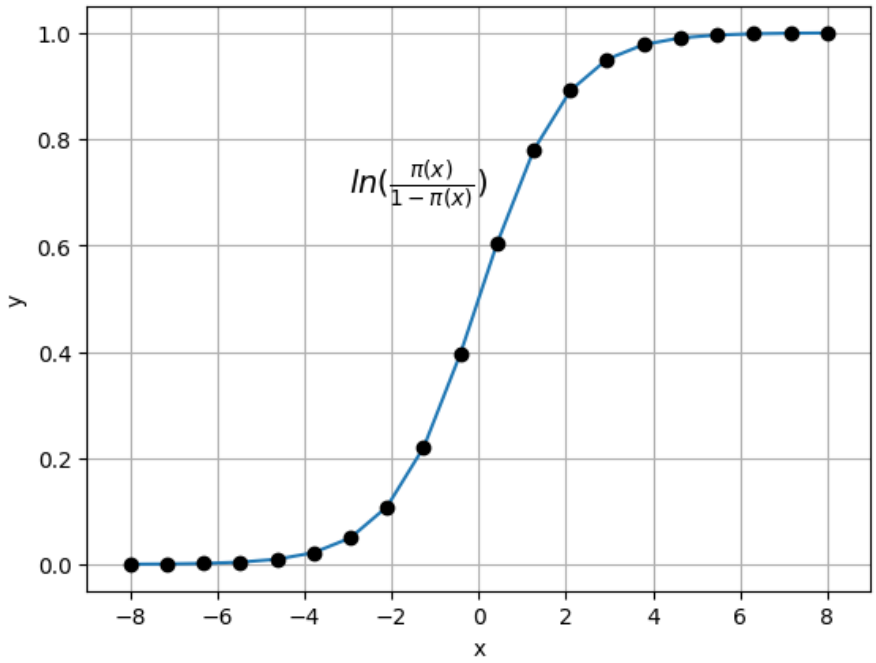


Fuente: elaboración propia con *Python*

De esta manera, la transformación *logit* expresada en la ecuación 12 previa que propone el modelo de regresión logística aplicada a una variable discreta binaria, permite estimar las

proporciones de Y a través de una transformación continua ln $\left(\frac{\pi(X)}{1-\pi(X)}\right)$. Gráficamente (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "1. Representación del proceso logístico"), este proceso arroja como resultado:

Gráfico 11: representación de la transformación *logit*



Fuente: elaboración propia con *Python*

A partir del modelo presentado, para poder medir como contribuye la responsabilidad en la construcción de privacidad de datos, se propone el siguiente modelo logístico, tomando las variables creadas en el apartado 4.2.2.

$$Privacidad = \beta_0 + \beta_1 * Epsilon + \beta_2 * Error + \beta_3 * N + \beta_4 * Responsabilidad \quad (14)$$

donde $Privacidad$ surgirá como resultado de la transformación ln $\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right)$, con $1 \leq i \leq 4$ ya que se posee 4 variables predictoras.

De este modo, para una evaluación preliminar de resultados en base a los datos con los cuales se cuenta, el objetivo es poner a prueba el modelo propuesto en los diferentes escenarios planteados en el apartado 4.2.2. Esto permitirá obtener el valor de los coeficientes que representan la contribución de cada predictora en la obtención de privacidad. En el siguiente apartado se lleva a cabo la corrida de los modelos y se valúan los resultados obtenidos.

4.3.2 **Aplicación del modelo y análisis de resultados**

A partir de las variables Privacidad, Épsilon, Error, Tamaño de muestra (N) y Responsabilidad, creadas y presentadas en el apartado 4.2.2 se realizan diferentes corridas, del modelo propuesto en el apartado anterior dado por la ecuación (14), con *Python* (Ver código en Apéndice – sección "Apéndice A3 Modelización de la Privacidad" – "2. Modelización de la privacidad y responsabilidad") en los diferentes escenarios propuestos. Se agrega el Escenario 0 como modelo base siendo aquel que no incorpora a la variable Responsabilidad. A continuación, se presenta una tabla resumen con los valores de los coeficientes obtenidos:

Tabla 10: coeficientes estimados por escenario

Coeficientes	β_0	β_1	β_2	β_3	β_4
Escenario 0	20.59108216	-26.869919	-2.13778694	-0.00523952	-
Escenario 1	4.76352056	-19.2974147	-1.51327234	-0.00369775	**10.83176774**
Escenario 2	3.33465727	-29.6922623	-1.1947045	-0.0004485	**8.17900757**
Escenario 3	8.80982703	-75.7389287	-2.02532143	0.00162966	**9.07816414**

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

En la Tabla 10, lo primero que es posible observar es que el signo que poseen los coeficientes β_1 a β_4 se corresponde con el obtenido en los valores de la covarianza presentados en el aparatado 4.2.3. De esta manera, el sentido de la contribución para obtener privacidad de cada variable predictora se mantiene. En segundo lugar, se observa principalmente la magnitud de contribución positiva de la Responsabilidad a través del valor de β_4. En el caso del Escenario 1, la contribución es de aproximadamente 11 veces por cada aumento en el cumplimiento de Responsabilidad; en el escenario 2 de 8 veces y en el escenario 3 de 9 veces. En comparación con el tamaño de muestra y el error, es muy superior independientemente del sentido de la contribución. Si se la compara respecto a Épsilon, es algo menor, pero este contribuye negativamente mientras que puede ser compensado por la contribución positiva de la Responsabilidad. No obstante, en el conjunto de las contribuciones, surge que la Responsabilidad es un elemento clave para la construcción de la privacidad de datos.

Para asegurar la conclusión anteriormente obtenida sobre la Responsabilidad, se evalúa la significatividad estadística de β_4 en el modelo. Para ello se utiliza el Test de Wald (De Jong y Heller, 2008; Agresti, 2012; Hosmer et al., 2013). Se plantean las siguientes hipótesis:

$$\begin{cases} H0: \beta_4 = 0 \ (no\ es\ siginificativo\ estadísticamente) \\ H0: \beta_4 \neq 0 \ (es\ siginificativo\ estadísticamente) \end{cases}$$

Criterio de decisión para rechazar H0: p-valores < alfa=0.05

Al realizar el Test con *Python,* se obtienen los siguientes p-valores:

Tabla 11: p-valor para cada coeficiente de Responsabilidad por escenario

Coeficientes	β_4
Escenario 1	≈ 0,000
Escenario 2	≈ 0,000
Escenario 3	≈ 0,000

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

A partir de la tabla 11, se observa que en todos los casos se obtiene un p-valor muy cercano a cero, razón por la cual se rechaza la hipótesis nula (H0). De este modo, resulta que el coeficiente β_4 es estadísticamente significativo lo que da validez a la conclusión anteriormente obtenida. De esta manera, es posible decir que el cumplimiento de la Responsabilidad resulta en un elemento clave relevante para la obtención de Privacidad de datos.

Otra mirada para evaluar la importancia de considerar a la Responsabilidad en la construcción de Privacidad se puede obtener de observar el ajuste del modelo. Para determinar si el modelo presenta un buen ajuste o no, se utiliza una medida de exactitud (*Accuracy*) en la predicción. Se define como el porcentaje de casos correctamente clasificados respecto del total de casos (Wang y Fu, 2005). Su cálculo viene dado por:

$$Accuracy = \frac{VP+VN}{VP+FP+VN+FN} \ (15)$$

donde VP (verdaderos positivos) son la cantidad de casos clasificados correctamente como positivos, VN (verdaderos negativos) son la cantidad de casos clasificados correctamente como negativos, FP (falsos positivos) son la cantidad de casos erróneamente clasificados como positivos y FN (falsos negativos) son la cantidad de casos erróneamente clasificados como negativos.

Realizando el cálculo con Python, se obtiene:

Tabla 12: Exactitud de la predicción de Privacidad por el modelo

	Accuracy
Escenario 0	0.9298
Escenario 1	0.9433
Escenario 2	0.9801
Escenario 3	0.9918

Fuente: elaboración propia a partir de los resultados obtenidos con *Python*

En base a los resultados mostrados en la tabla 12, es posible determinar que la incorporación de la variable Responsabilidad (escenario 1, escenario 2 y escenario 3 en comparación con el escenario 0 que no contempla a la variable responsabilidad), contribuye a mejorar la exactitud de predicción realizada por el modelo. Esto permite revalidar la importancia de considerar a la responsabilidad organizacional en la construcción de Privacidad de datos.

A partir del análisis de los resultados obtenidos de la aplicación del modelo, es posible decir que la responsabilidad organizacional resulta en un elemento clave para brindar protección de datos personales como garantía de privacidad de datos. Si bien este resultado era esperado dado el planteo de discusión desarrollado en el marco teórico, resulta en un punto de partida para ser considerado a la hora de diseñar políticas en base a datos personales en organizaciones públicas. En particular, se expone la necesidad de promover (y de revisión apremiante de cualquier regulación vigente) una cultura organizacional de gestión de datos bajo un marco de responsabilidad. Además, surge que un enfoque cuantitativo resulta relevante en como herramienta de gestión de datos personales.

Conclusión de capítulo

En el presente capítulo, se realizó la modelización logística de la privacidad y su aplicación para ciertos valores con el fin de determinar el grado de contribución de la responsabilidad organizacional en la construcción de privacidad de datos personales. El principal resultado obtenido en el contexto de aplicación realizada es que la responsabilidad contribuye significativamente en la obtención de privacidad. Esto permitió validar parcialmente lo planteado en la discusión teórica desarrollada a lo largo de esta tesis, pero mostrando el potencial herramental de lo realizado.

Para llevarlo a cabo, se operacionalizaron los elementos seleccionados como claves en la obtención de privacidad de datos mediante la construcción de variables cuantitativas. Partiendo de una definición inicial de ciertas reglas (que podrían ser ampliadas) para la construcción de la variable Privacidad, se presentaron diferentes escenarios para alcanzar cierta generalización de resultados. El aporte central de este capítulo en particular, y de la tesis en general, ha sido la determinación del grado de relevancia de la responsabilidad organizacional en términos de protección de datos personales para brindar garantías de privacidad a los ciudadanos. Para lograr esta contribución, el capítulo se ha dividido en tres apartados.

Dentro del primero, se establecieron los aspectos principales a ser tenidos en cuenta para llevar adelante el diseño de una estrategia de gestión responsable de datos personales por defecto y desde el diseño. Estos se encuadran dentro de lo desarrollado en el capítulo 1 y 2, y son las organizaciones e individuos, la tecnología, los procesos de datos y las políticas y principios. Una vez establecidos los principios que aportan las definiciones conceptuales para el diseño, se elaboró una propuesta de plan estratégico para la gestión de la privacidad.

Para poder llevar adelante el plan propuesto, se identificaron las actividades principales y específicas que permiten alcanzar el control del riesgo de privacidad asociada a datos personales. Estas implican tareas operativas en diferentes etapas del procesamiento de datos para actuar de manera preventiva antes que reactiva. Como resultado, se logra establecer un marco de acción en un contexto organizacional en línea con la definición de privacidad de datos personales propuesta en esta tesis. A su vez, la obtención de resultados concretos permitió su cuantificación para realizar una evaluación. De este modo, se habilita la posibilidad de aplicar una técnica cuantitativa para medir el grado de asociación entre la responsabilidad organizacional y la construcción de privacidad de datos.

Para poder alcanzar la medición del grado de asociación entre responsabilidad y privacidad, en un segundo apartado, se realiza la selección de los elementos claves. Esta selección se basó, por un lado, en lo desarrollado en el capítulo 3 mediante la selección y aplicación de la metodología de la Privacidad Diferencial a un caso para la protección de datos personales de donde se obtuvieron 3 elementos claves (Épsilon, Error y Tamaño de muestra). Por el otro, mediante la creación de una variable denominada Responsabilidad cuyos valores posibles reflejan el nivel de cumplimiento de las actividades previamente definidas.

Finalmente, se construye la variable Privacidad que puede tomar valor 1 (obtención de privacidad) o 0 en base a ciertas reglas en diferentes escenarios.

Al definirse a la variable Privacidad como binaria ya que solo puede tomar valor 0 o 1, se propuso el método de la regresión logística para modelizar adecuadamente el vínculo entre las variables predictoras (Épsilon, Error, Tamaño de muestra y Responsabilidad) y la variable objetivo (Privacidad). Dado que este modelo predice mediante un vínculo lineal, previamente de realizó un análisis estadístico sobre la existencia de este. Como resultado, en los diferentes escenarios planteados, se pudo establecer que existe relación lineal de manera significativa entre cada predictora y la variable objetivo.

Finalmente, en el tercer apartado, se aplica el modelo logístico propuesto en los diferentes escenarios y se analizan los resultados obtenidos. El principal resultado alcanzado es la detección de como la variable Responsabilidad contribuye significativamente en la obtención de Privacidad. El nivel de cumplimiento de las tareas específicas (Responsabilidad operativa), contribuye en hasta 11 veces en la obtención de Privacidad. Si bien las evaluaciones realizadas no abarcan la totalidad de posibilidades, en todos los casos los resultados son favorable. Esto permite interpretar que la Responsabilidad organizacional es un elemento clave para alcanzar garantías de privacidad.

Tener en cuenta estos resultados preliminares, permite aproximar a concluir que la responsabilidad es necesario sea incorporada para llevar adelante una gestión responsable de datos personales en organizacionales públicas, aunque también en privadas. Si bien este resultado era esperado, debido a que es una discusión que viene llevándose a cabo en los últimos años desde diferentes perspectivas, también se constituye como punto de partida para la realización de políticas públicas basadas en información. En particular, este resultado preliminar deja expuesta la necesidad de incluir a la responsabilidad organizacional dentro de la regulación argentina. Además, que sea adoptada como parte integral de la cultura organizacional independientemente de la anterior, de forma tal de contribuir en el desarrollo del bienestar social. No solo alcanza con que las diferentes agencias estatales impulsen los principios en torno a la problemática, sino que resulta necesario adoptar un accionar preventivo en términos operativos.

Por otra parte, la detección y posterior operacionalización de los elementos claves expuesto a lo largo del capítulo, permite contar con un enfoque cuantitativo para emplear técnicas que

se constituyen en herramientas de gestión para las organizaciones. Teniendo en cuenta la detección de riesgos realizada en un capítulo anterior, todo lo expuesto puede ser considerado como un punto de partida para la elaboración de una metodología estandarizada. En este sentido, constituye un aporte fundamental para impulsar una gestión de datos personales que permita incorporar un criterio basado en responsabilidad para el accionar organizacional.

Conclusión

El principal aporte de la tesis es la propuesta de un modelo de gestión de la privacidad de datos considerando a la responsabilidad organizacional como elemento clave al contemplar la dinámica del mercado de datos y la regulación. Su aplicación no solo expone la importancia de contar con un estándar metodológico que sea transparente y explicable para la ciudadanía, sino que brinda garantías de confianza para fortalecer el desarrollo de políticas públicas en base a datos. En este sentido, se constituye en una herramienta de gestión bajo un marco de responsabilidad para las organizaciones, en particular las públicas del Estado, con el fin de contribuir al bienestar de la sociedad. Se espera que el diseño propuesto centrado en la responsabilidad sea adoptado por las organizaciones al momento de diseñar un marco de gestión para el tratamiento de datos personales.

El objetivo general que se propuso en esta tesis es la determinación de los elementos estructurales claves para el diseño de una estrategia de gestión responsable de la privacidad de datos personales en organizaciones estatales argentinas. Su cumplimiento parcial, fue posible gracias a la construcción de una base datos personales que luego de ser procesada y analizada, los resultados obtenidos permitieron la construcción de variables cuantitativas para ser modelizadas. La relevancia de tal análisis se basa en considerar que la recopilación y uso de datos personales por parte de las organizaciones públicas conlleva riesgos que son co-constituidos entre estas, individuos y tecnología. Por esta razón, la responsabilidad organizacional emerge como variable clave frente a los marcos de regulación existentes, así como también para asegurar el desarrollo de políticas públicas basadas en datos que beneficie a la sociedad en conjunto. Incluso, al considerar el veloz avance de la capacidad tecnológica para la captura y obtención de datos, la ausencia de tal marco centrado en la responsabilidad puede generar un obstáculo para futuros desarrollos.

La realización de la tesis se organizó en cuatro capítulos, agrupados en dos partes. Los capítulos 1 y 2, abarcaron la presentación del corpus teórico. En el primero se desarrolló sobre la conformación del mercado de datos y el riesgo derivado para la privacidad de datos. Se explicó como la utilización de tecnología para el procesamiento de datos por parte de las organizaciones genera una asimetría de poder que vulnera el derecho a la privacidad de los individuos. A partir de ello, en el segundo capítulo se argumenta porque se requiere incorporar a la responsabilidad organizacional para brindar garantías de protección de datos personales a los individuos. Los capítulos tres y cuatro, presentaron los modelos desde un

enfoque cuantitativo. En el capítulo 3, se eligió un método mediante el diseño de un algoritmo para alcanzar la protección de datos personales mediante distorsión sin invalidar su posterior uso. A partir de ello, surgen los tres primeros elementos claves para el diseño de una estrategia responsable de privacidad de datos, abordada en el capítulo 4. En este último, se sumó la creación de una variable para cuantificar el grado de responsabilidad proveniente del cumplimiento de tareas operativas en relación con principios fundacionales de la privacidad de datos. Además, se creó una variable objetivo que representa si se alcanza privacidad o no. De este modo, fue posible modelizar y analizar el grado de asociación entre privacidad y responsabilidad.

El objetivo del primer capítulo fue analizar cómo se conforma un nuevo mercado de datos a partir de la interacción entre individuos, tecnología y organizaciones que puede vulnerar la privacidad de los primeros. La hipótesis asociada a este capítulo fue que el procesamiento de datos personales por parte de las organizaciones públicas estatales pone en riesgo la privacidad de los individuos. A través del desarrollo llevado a cabo, se explicó como el mercado de datos se co-constituye a partir de la interacción en red de los actores intervinientes. A su vez, se argumentó como surge un riesgo de violación de la privacidad en base a la utilización de datos personales mediante un proceso omnipresente creado por las organizaciones en este contexto.

El principal aporte de este primer capítulo fue argumentar la existencia de una asimetría de información que surge del procesamiento de datos a través de tecnología realizado por las organizaciones. A su vez, esto da origen y sustento a una asimetría de poder de control por parte de las organizaciones públicas del Estado. Si bien, tal capacidad de acción facilita la generación de conocimiento para la elaboración de políticas públicas eficientes y eficaces, también puede ser utilizada para modificar el escenario de acciones posibles del individuo. A su vez, como los datos personales son utilizados y en muchos casos no protegidos o poco protegidos, incluso pueden incurrir en la violación del derecho a la privacidad de los ciudadanos. De aquí que la responsabilidad organizacional emerge como relevante en este escenario.

El objetivo del segundo capítulo fue elaborar una definición de la privacidad de datos personales que incorpore a la responsabilidad organizacional como factor clave para la construcción de privacidad desde el diseño y por defecto. La hipótesis planteada fue que la responsabilidad organizacional para abordar la problemática de la privacidad de datos

personales en términos operativos es necesaria para mitigar el riesgo asociado. Mediante la incorporación de la perspectiva regulatoria europea y argentina – en contra posición, a la norteamericana- que considera a la privacidad como un derecho humano, se amplía y consolida la argumentación planteada en el capítulo 1. Si bien es necesaria la regulación para una convivencia en sociedad, aun así, su carácter reactivo resulta insuficiente para brindar garantías de privacidad de datos personales. A partir de ello, la responsabilidad organizacional se posiciona como necesaria para llevar adelante una gestión responsable.

A partir de la conclusión anterior, el principal aporte del capítulo es la elaboración de una definición de la privacidad de datos personales que incluye a la responsabilidad organizacional. Se la definió como la posibilidad de que un individuo generador de datos al interactuar con una organización en un contexto de red digital pueda obtener garantías suficientes de que sus datos están siendo debidamente protegidos. Por un lado, esta definición reconoce el derecho humano a la privacidad que poseen los individuos permitiendo que puedan decidir si aceptan o no que una organización procese sus datos personales. Por el otro, implica que las organizaciones públicas y privadas deben asumir una gestión responsable que garantice la protección de los datos personales que recopilan y utilizan con el fin de brindar garantías de privacidad a los ciudadanos.

Contemplar la definición propuesta, si bien requiere de un consenso desde diferentes enfoques disciplinares, permite incorporar a la responsabilidad organizacional en términos operativos a la hora de utilizar datos personales. Esto, se encuentra en línea con el enfoque de la privacidad desde el diseño y por defecto, mediante el cual se adoptan ciertos principios que guían el accionar operativo de una organización para la construcción de privacidad. Su objetividad permite direccionar hacia la selección e implementación de alguna técnica concreta que brinde garantías de protección de datos personales, es decir, ir en búsqueda de mitigar el riesgo de privacidad asociado. Por esta razón, se propone la utilización de la metodología de Privacidad Diferencial. Si bien esto aborda parcialmente la propuesta definitoria realizada al ser solo un enfoque cuantitativo, permite considerarlo como una primera herramienta de gestión responsable en términos operativos.

Advirtiendo dicha limitación, se propone continuar utilizando el abordaje operativo con un enfoque cuantitativo para lograr construir un desarrollo de aplicación de la Privacidad Diferencial. Por un lado, para poder exponer su potencial para ser recomendada como una herramienta metodológica a ser adoptada por las organizaciones tanto públicas como

privadas. También, como un aporte a la comunidad académica de las ciencias económicas ya que se trata de vincular datos, tecnología y fundamentos estadísticos matemáticos en un contexto donde cada vez más las decisiones se toman basadas en información. Por el otro, para poder detectar ciertos elementos que se posicionaran como claves en pos de la construcción de la responsabilidad organizacional en tanto necesaria para el diseño de una estrategia de gestión responsable de la privacidad.

La segunda parte de la tesis, donde se desarrolla el enfoque cuantitativo, comienza con el capítulo 3. El objetivo fue seleccionar una metodología cuantitativa para impedir la identificación de individuos en base a datos sensibles que otorga garantía de privacidad a los ciudadanos. La hipótesis establecida en este capítulo es que el método de la Privacidad Diferencial es efectivo para la protección de datos personales. Mediante la aplicación de caso, se alcanza su validación parcial. El hecho de utilizar una única base de datos expone la limitación del alcance de la aplicación realizada. No obstante, como la muestra utilizada fue construida en base a datos personales reales sobre 1000 ciudadanos argentinos, los resultados obtenidos se consideran suficientes para mostrar su potencialidad.

La Privacidad Diferencial es un método matemático probabilístico que permite aplicar ruido aleatorio a los datos para brindar cierta garantía de privacidad. Pero también resulta importante que permite encontrar un equilibrio entre la necesidad de uso de los datos y su protección como herramienta de control. Esta decisión puede ser abordada a partir de su parámetro clave que es Épsilon, ya que brinda una medida acerca de la pérdida de privacidad en la que se incurre para diferentes niveles de ruido aleatorio aplicado. Para llevar adelante la aplicación mostrada, los datos utilizados fueron recopilados manualmente a partir de tomar una muestra de 1000 CUIT (Clave Única de Identificación Tributaria). Con cada uno estos se consultaron diferentes servicios *web* de organizaciones públicas para la obtención de otros atributos personales. Finalizada su construcción, mediante la utilización del lenguaje de programación *Python*, se llevó a cabo la aplicación del método mediante el desarrollo de un algoritmo. Posteriormente se realizó una evaluación de resultados.

La aplicación realizada permite considerar el potencial que ofrece el método de la Privacidad Diferencial como una herramienta operativa para llevar adelante un accionar responsable sobre datos personales en un contexto organizacional. Las tareas involucradas comenzaron con la detección de riesgos de identificación de individuos en la base de datos utilizada, lo que constituye una primera acción de control. La posterior aplicación de la mencionada

metodología arrojó como principal resultado el romper con la posibilidad de que un usuario de los datos logre identificar a ciudadanos contenidos en la base trabajada. Por esta razón se concluye que el método propuesto tiene un potencial protector de datos personales, que puede ser utilizado para brindar garantías de privacidad y ser empleada en contextos organizacionales.

En el último capítulo de la tesis, se propuso como objetivo la selección de los elementos estructurales claves para llevar adelante el diseño de una estrategia de gestión responsable de datos personales en organizaciones. Si bien el capítulo no posee una hipótesis especifica, se deriva como consecuencia de los desarrollado previamente. En este sentido se utiliza un enfoque interpretativo para la determinación de los elementos claves que surgen de los resultados previos alcanzados. Luego, mediante un enfoque cuantitativo se operativizaron dichos elementos mediante la creación de variables cuantitativas. Esto permitió proponer un modelo para representar el vínculo entre privacidad y responsabilidad y poder evaluar el grado de contribución de esta última en la construcción de la primera.

Retomando el enfoque de privacidad desde el diseño y por defecto presentado en el marco teórico, se especificaron cuáles son las tareas operativas que conducen al cumplimiento de los principios fundacionales de privacidad enunciados. A partir de ello, se elabora una propuesta de diseño de una estrategia de gestión responsable de la privacidad. Bajo este paragua, se determina que los elementos claves son: Épsilon, Error, Tamaño de muestra y la Responsabilidad.

En primer lugar, se realizó nuevamente una corrida del algoritmo de Privacidad Diferencial desarrollado sobre la misma muestra de 1000 datos personales, pero con mayor cantidad de valores de los parámetros para representar la mayor cantidad de posibilidades. Como resultado, se obtienen un listado de valores de los parámetros Épsilon, Error, Tamaño de muestra. A partir de ello se los convirtió en variables cuantitativas. En segundo lugar, se construye la variable Responsabilidad. Para ello se seleccionaron determinados valores que emulan diferentes niveles porcentuales de cumplimiento del conjunto de actividades operativas descriptas previamente. Realizando cierta simplificación, esta quedó conformada con los valores 0, 0.25, 0.50, 0.75 y 1 para emular porcentajes de cumplimiento. Finalmente, se construye la variable Privacidad que puede tomar valor 0 (no se logra) y 1 (se logra), considerando diferentes escenarios mediante la construcción de reglas simplificadas para la definición del valor 1. Si bien, no fue posible abordar las infinitas posibilidades de escenarios

que pueden presentarse reconociendo la limitación del alcance de la propuesta, los resultados alcanzados posteriormente permitieron realizar una interpretación del grado de contribución de la Responsabilidad en la construcción de la privacidad.

A través de la aplicación de un modelo de regresión logística que permite vincular linealmente la relación entre variables predictoras y una variable objetivo-binaria, se aproximaron los primeros resultados. Previamente, para saber si existía relación lineal entre las variables definidas anteriormente, se calculó la correlación y se evaluó la significatividad estadística de esta. Como el resultado de este análisis estadístico inicial fue favorable, se llevó a cabo la aplicación del modelo logístico propuesto en los diferentes escenarios construidos. A partir de los resultados obtenidos pudo interpretarse que la Responsabilidad contribuye en hasta 11 veces en la obtención de Privacidad. Esto permitió, por un lado, medir operativamente la contribución de esta variable, a pesar de la cantidad limitada de escenarios evaluados. Por el otro, exponer la importancia de contar con herramientas cuantitativas para llevar adelante un accionar responsable dentro de un contexto organizacional.

Este resultado preliminar es de especial interés para el desarrollo de una gestión responsable de la privacidad de datos personales. El mismo indica que existen indicios de que la responsabilidad organizacional es un elemento clave para ser incorporado como criterio a la hora de diseñar una estrategia de gestión en organizaciones públicas. Este criterio se encuentra asociado a llevar adelante acciones que contribuyan al bienestar social. Como se mencionó en el corpus teórico, ante la insuficiencia de la normativa vigente en Argentina por su carácter reactivo, el accionar organizacional preventivo bajo un marco de responsabilidad se constituye en fundamental.

El análisis en general y la construcción de las variables vinculadas a través de un modelo, son un primer paso para direccionar hacia la gestión responsable de datos personales en organizaciones públicas del estado argentino. La sola identificación de los elementos claves que en términos operativos permiten evaluar un accionar más o menos responsable, plantean un escenario de reflexión y potenciales resultados. Además, como fue mencionado en el capítulo 2, poder contar con una metodología estandariza que evalúe el grado de contribución a la construcción de privacidad de datos, aporta a la transparencia de los procesos aplicados de cara a la sociedad. En este sentido, se espera por un lado que las organizaciones, puedan acceder a implementar lo propuesto. Además, ante la posibilidad de una revisión para la actualización de la normativa vigente, también pueda ser contemplado.

Por otra parte, el aporte de los desarrollos realizados facilita la transferencia de los resultados de la tesis a diferentes actores. Para el sector público, se pone a disposición el trabajo realizado, brindando una herramienta para gestionar responsablemente los datos personales que recopilan y utilizan para la elaboración de políticas públicas. En cuanto al sector científico académico, el trabajo provee un enfoque operativo para llevar adelante la elaboración de tareas de gestión bajo un marco de responsabilidad. A su vez, esto permite impulsar un perfil de formación profesional en un sentido integral bajo una perspectiva ética.

En particular para el sector académico, el trabajo realizado puede ser utilizado como una fuente de consulta para la realización de futuros trabajos que utilicen datos personales bajo un marco de responsabilidad. Además, la interdisciplinaridad involucrada en la temática trabajada invita a generar nuevos espacios de discusión en comunión académica. La generación de nuevos conocimientos en un contexto de cambio profundo por el impacto de la tecnología invita a la reflexión contínua para seguir trabajando en pos del bienestar social. Con respecto a la metodología utilizada dentro del trabajo, también aporta un proceder con razonamiento lógico para abordar una problemática tan sensible en la actualidad. En este sentido, la tesis se constituye como un material de estudio en el ecosistema de datos dentro de la articulación académica entre doctorado, maestría y grado.

La tesis presenta diversas posibilidades para ampliar su alcance. En primer lugar, porque se realizó un análisis a partir de tomar un listado de CUIT publicados por una organización de gobierno de la Ciudad de Buenos Aires, pero no se hizo ninguna propuesta concreta para que sea incorporada en esta. En segundo lugar, es posible expandir el alcance del estudio realizado. Por ejemplo, ampliando la cantidad de datos utilizados o trabajando en conjunto en base a los datos que poseen una o varias organizaciones públicas. En tercer lugar, desarrollar un aplicativo que agilice la aplicación realizada al mismo tiempo que impida el acceso directo a los datos personales por parte del usuario. En este sentido, el desarrollo de un aplicativo se presenta como una nueva oportunidad a la vez que permitiría generalizar y estandarizar una metodología de trabajo.

Se considera de suma importancia que el enfoque responsable para la gestión de datos personales sea incorporado con cierto apremio. Esto porque los avances tecnológicos para el procesamiento de datos evolucionan a pasos agigantados día tras día. Por esta razón, para aprovechar al máximo las potencialidades que estas nuevas innovaciones brindan, resulta fundamental anteponer el accionar responsable junto con la reflexión conjunta en comunidad. De este modo, se podrá

ampliar el accionar transparente del Estado en conjunto con los ciudadanos para contribuir a un desarrollo social responsable en el país.

Referencias bibliográficas

Aboody, D., & Lev, B. (2000). Information asymmetry, R&D, and insider gains. The journal of Finance, 55(6), 2747-2766. DOI: 10.1111/0022-1082.00305

AEPD (Agencia Española de Protección de Datos) [en línea] https://www.aepd.es/es/guias-y-herramientas/guias

——— (2019), Guía de Privacidad desde el Diseño. [en línea] https://www.aepd.es/es/documento/guia-privacidad-desde-diseno.pdf

——— (2020), Guía de Protección de Datos por Defecto. [en línea] https://www.aepd.es/sites/default/files/2020-10/guia-proteccion-datos-por-defecto.pdf

Agresti, A. (2012). Categorical data analysis (Vol. 792). John Wiley & Sons. ISBN:9780471360933

ambito.com, (23 de abril de 2023). Denuncian penalmente a la Ciudad por el uso de datos biométricos sin "justificación racional". *ámbito*. https://www.ambito.com/politica/denuncian-penalmente-la-ciudad-el-uso-datos-biometricos-justificacion-racional-n5705668

Aoun Barakat, K. & Sayegh, M. (2021). Information Asymmetry in the Age of Big Data Analytics. Conference paper: Cognitive Analytics Management Conference, Feb 2021 at Beirut.

Arner, D. W., Castellano, G. G., & Selga, E. (2022). Financial Data Governance: The Datafication of Finance, the Rise of Open Banking and the End of the Data Centralization Paradigm. University of Hong Kong Faculty of Law Research Paper, (2022/08)

Aykut, S. C., Demortain, D., & Benbouzid, B. (2019). The Politics of Anticipatory Expertise: Plurality and Contestation of Futures Knowledge in Governance â Introduction to the Special Issue. Science & Technology Studies, 32(4), 2-12.

Bell, D., (1979). The Social Framework of the Information Society. In Dertouzos, M. L., & Moses, J. (Eds), The Computer Age, (pp. 163-211). The MIT Press, Cambridge, Massachusetts, and London, England.

Bryson, J. M., Crosby, B. C., & Bloomberg, L., (2015). Creating public value in practice: Advancing the common good in a multi-sector, shared-power, no-one-wholly-in-charge world. CRC Press. ISBN 9781482214604

Çalışkan, K., & Callon, M. (2010). Economization, part 2: a research programme for the study of markets. Economy and society, 39(1), 1-32.

Çalışkan, K., & Callon, M. (2009). Economization, part 1: shifting attention from the economy towards processes of economization. Economy and society, 38(3), 369-398.

Callon, M. & Muniesa, F. (2003). Les marchés économiques comme dispositifs collectifs de calcul. Réseaux, vol. no 122, no. 6, 2003, pp. 189-233. https://www.cairn.info/revue--2003-6-page-189.htm.

Canavos, G. C., & Medal, E. G. U. (1987). Probabilidad y estadística (p. 651). México: McGraw Hill.

Castells, M. (2002). La era de la información. Vol. I: La Sociedad Red. Siglo XXI Editores. México, Distrito Federal.

Cavoukian, A. (2011), Privacy by Design: The 7 Foundational Principles. Ph.D. Information & Privacy Commissioner Ontario, Canadá. https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

CENIL (Commission Nationale de l'Informatique et des Libertés). Data protection around the world. [en línea] https://www.cnil.fr/en/data-protection-around-the-world

Clarke, R. (1999). Internet privacy concerns confirm the case for intervention. Communications of the ACM, Volume 42, Issue 2, Feb. 1999, pp 60–67 https://doi.org/10.1145/293411.293475

Cleven, A., & Wortmann, F. (2010, January). Uncovering four strategies to approach master data management. In 2010 43rd Hawaii International Conference on System Sciences (pp. 1-10). IEEE. DOI: https://doi.org/10.1109/HICSS.2010.488

Constantiou, I., Kallinikos, J. (2015) New games, new rules: big data and the changing context of strategy. J Inf Technol 30, 44–57. https://doi.org/10.1057/jit.2014.17

Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R., & Haddadi, H. (2016). Enabling the new economic actor: data protection, the digital economy, and the Databox. Personal and Ubiquitous Computing, 20, 947-957.

Culpepper, P. D., & Thelen, K. (2020). Are we all Amazon primed? Consumers and the politics of platform power. Comparative Political Studies, 53(2), 288-318. https://doi.org/10.1177/0010414019852687

Custers, B. (2013). Data Dilemmas in the Information Society: Introduction and Overview. In Calders, T., Schermer, B., Zarsky, T., & Custers, B. (Eds.), Discrimination and Privacy in the Information Society (pp. 3-26). Springer. ISBN: 978-3-642-30487-3

Cutolo, D., & Kenney, M. (2021). Platform-dependent entrepreneurs: Power asymmetries, risks, and strategies in the platform economy. Academy of Management Perspectives, 35(4), 584-605.

De Jong, P., & Heller, G. Z. (2008). Generalized linear models for insurance data. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511755408

Dwork, C. (2008). *Differential Privacy: A Survey of Results*. In: Agrawal, M., Du, D., Duan, Z., Li, A. (eds) Theory and Applications of Models of Computation. TAMC 2008. Lecture Notes in Computer Science, vol 4978. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-79228-4_1

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci., 9(3-4), 211-407. https://doi.org/10.1561/0400000042

Eberendu, A. C. (2016). Unstructured Data: an overview of the data of Big Data. International Journal of Computer Trends and Technology, 38(1), 46-50.

Erickson, J. (1999). Algorithms. Creative Commons Attribution 4.0 International license. https://jeffe.cs.illinois.edu/teaching/algorithms/book/Algorithms-JeffE.pdf

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. Early engagement and new technologies: Opening up the laboratory, 55-95.

Gandy Jr, Oscar H. (2021). Information and Power. In O. Gandy Jr, The panoptic sort: A political economy of personal information (pp. 29-50). Oxford University Press, 2021. https://doi.org/10.1093/oso/9780197579411.001.0001

García Fronti, J., & Matías Herrera, P. (2021). Mercado de datos personales: asimetrías entre plataformas e individuos. Gestión Joven, 22(3).

García Jiménez, E., Gil Flores, J., & Rodríguez Gómez, G. (1994). Análisis de datos cualitativos en la investigación sobre la diferenciación educativa. Revista de investigación educativa, 23, 179-213. URI: https://hdl.handle.net/11441/77867

Guston, D. H. (2014). Understanding 'anticipatory governance'. Social studies of science, 44(2), 218-242.

Haskel, J., & Westlake, S. (2017). Capitalism without capital. Princeton University Press. ISBN: 9780691175034

Himma, K. E., & Tavani, H. T. (Eds.). (2008). The handbook of information and computer ethics. Hoboken: Wiley & Sons, Inc. DOI:10.1002/9780470281819

Hoepman, J. H. (2022). Privacy design strategies (the little blue book). https://www.cs.ru.nl/~jhh/publications/pds-booklet.pdf

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons. ISBN:9780470582473

HromkoviČ, J. (2005). *Fundamentals*. In HromkoviČ, J, Design and Analysis of Randomized Algorithms. Introduction to Design Paradigms (pp. 19-99). Springer Berlin, Heidelberg. ISBN: 978-3-540-27903-7

Jurafsky, D., & Martin, J. H. (2024) Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. In Standfor University Web: https://web.stanford.edu/~jurafsky/slp3/5.pdf

Kanakia, H., Shenoy, G., Shah J. (2019). Cambridge Analytica – A Case Study. Indian Journal of Science and Technology, Vol 12(29), DOI: 10.17485/ijst/2019/v12i29/146977

Kolanovic, M., & Krishnamachari, R. T. (2017). Big data and AI strategies: Machine learning and alternative data approach to investing. JP Morgan Global Quantitative & Derivatives Strategy Report. https://cpb-us-e2.wpmucdn.com/faculty.sites.uci.edu/dist/2/51/files/2018/05/JPM-2017-MachineLearningInvestments.pdf

Kshetri, N. (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy, 38(11), 1134-1145. DOI: 10.1016/j.telpol.2014.10.002

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Harvard university press. ISBN 9780674792913

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford university press. ISBN: 9780199256051

Latour, B. (2011). Network Theory| Networks, Societies, Spheres: Reflections of an Actor-network Theorist. International Journal Of Communication, 5, 15. Retrieved from https://ijoc.org/index.php/ijoc/article/view/1094/558

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (2009). A brief history of the Internet. ACM SIGCOMM computer communication review, 39(5), 22-31.

Mattelart, A., (2001). *Histoire de la Société de l'information*. Éditions La Découverte, París, Francia. Edición en español (2002), Ed. Paidós.

Mazzucato, M., Entsminger, J., & Kattel, R. (2020). Public Value and Platform Governance. UCL Institute for Innovation and Public Purpose WP 2020-11. http://dx.doi.org/10.2139/ssrn.3741641

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68. https://hbr.org/2012/10/big-data-the-management-revolution

McKeen, J., & Smith, H. (2007). Developments in Practice XXIV: Information Management: The Nexus of Business and IT. Communications of the Association for information System. DOI 10.17705/1CAIS.01903

Milgrom, P. (1992). Economics, organization and management. Prentice-Hall, Inc. ISBN: 9780132246507

Moazed, A., & Johnson, N. L. (2016). Modern monopolies: what it takes to dominate the 21st century economy. St. Martin's Press. ISBN: 9781250091895

Mortier, R., & Haddadi, H., & Henderson, T., & McAuley, D., & Crowcroft, J. (2014). Human-Data Interaction: The Human Face of the Data-Driven Society DOI: http://dx.doi.org/10.2139/ssrn.2508051

Naser, A. (2021). Gobernanza digital e interoperabilidad gubernamental: una guía para su implementación. Documentos de Proyectos (LC/TS.2021/80), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL). URI https://hdl.handle.net/11362/47018

Near, J. P., & Abuah, C. (2021). Programming Differential Privacy. Vol 1. [en línea] https://uvm-plaid.github.io/programming-dp/

OCDE (Organización de Cooperación y Desarrollo Económicos) (2019), Índice de gobierno digital OCDE 2019: resultados y mensajes clave. [en línea] https://www.oecd.org/gov/recomendacion-del-consejo-sobre-gobierno-abierto-141217.pdf

——— (2014), Recommendation of the Council on Digital Government Strategies. [en línea] https://one.oecd.org/document/C(2014)88/en/pdf

Rhodes, T., & Lancaster, K. (2019). Evidence-making interventions in health: A conceptual framing. Social Science & Medicine, 238, 112488. https://doi.org/10.1016/j.socscimed.2019.112488

Rivoir, A. L., & Morales, M. J. (2019). Derechos digitales y regulación de Internet. Aspectos claves de la apropiación de tecnologías digitales. En A. L. Rivoir, & M. J. Morales (Eds.). Tecnologías digitales: Miradas críticas de la apropiación en América Latina (pp. 35-49). CLACSO. http://biblioteca.clacso.edu.ar/clacso/se/20191128031455/Tecnologias-digitales.pdf

Rubinstein, I. (2012). Big data: The end of privacy or a new beginning?. International Data Privacy Law (2013 Forthcoming), NYU School of Law, Public Law Research Paper, (12-56). http://dx.doi.org/10.2139/ssrn.2157659

Rousseau, D. M. (2006). Is there such a thing as “evidence-based management”?. Academy of management review, 31(2), 256-269.

Salaberry, N. R., & Herrera, P. M. (2021). Gestión de la privacidad de datos sensibles.: El aplicativo CuidAR para el control del COVID-19. Gestión Joven, 22(2), 48-58. ISSN-e 1988-9011

Schmarzo, B. (2013). Big Data. Understanding How Data Powers Big Business. John Wiley & Sons, Inc. ISBN: 978-1-118-73957-0

Seidl, T. (2021). Commodification and Disruption: Theorizing Digital Capitalism. Centre for European Integration Research. WP, 02/2021. https://eif.univie.ac.at/workingpapers/index.php

Shapiro, C., Carl, S., & Varian, H. R. (1998). Information rules: a strategic guide to the network economy. Harvard Business Press. https://doi.org/10.2307/1183273

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. MIS quarterly, 989-1015. DOI: 10.2307/41409970

Steinmann, M., Matei, S. A., & Collmann, J. (2016). A theoretical framework for ethical reflection in big data research. Ethical reasoning in big data: An exploratory analysis, 11-27.

Stiglitz, J. E. (2000). The contributions of the economics of information to twentieth century economics. The quarterly journal of economics, 115(4), 1441-1478. https://doi.org/10.1162/003355300555015

Sweeney, L. (2000). Simple demographics often identify people uniquely. Health (San Francisco), 671(2000), 1-34.

Täuscher, K., & Laudien, S. M. (2018). Understanding platform business models: A mixed methods study of marketplaces. European Management Journal, 36(3), 319-329. https://doi.org/10.1016/j.emj.2017.06.005

UNCTAD (United Nations Conference on Trade and Development). Data Protection and Privacy Legislation Worldwide. [en línea] https://unctad.org/page/data-protection-and-privacy-legislation-worldwide

ur Rehman, I. (2019). Facebook-Cambridge Analytica data harvesting: What you need to know. In Library Philosophy and Practice, 1-11.

van de Waerdt, P. J. (2020). Information asymmetries: recognizing the limits of the GDPR on the data-driven market. Computer Law & Security Review, 38, 105436.

Varian, H. R. (2010). Computer mediated transactions. American Economic Review, 100(2), 1-10.

Verhulst, S. G. (2023). Operationalizing digital self-determination. Data & Policy, 5, e14. doi:10.1017/dap.2023.11

Victorelli, E. Z., Dos Reis, J. C., Hornung, H., & Prado, A. B. (2020). Understanding human-data interaction: Literature review and recommendations for design. International journal of human-computer studies, 134, 13-32. DOI: 10.1016/j.ijhcs.2019.09.004

Wang, L., & Fu, X. (2005). Data mining with computational intelligence. Springer Science & Business Media. ISBN : 978-3-540-24522-3

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. Journal of information technology, 30(1), 75-89. https://doi.org/10.1057/jit.2015.5

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Profile books. ISBN-13 9781610395694

Zwick, D., & Dholakia, N. (1999). Models of privacy in the digital age: Implications for marketing and e-commerce. Research Institute for Telecommunications and Information Marketing (RITIM), University of Rhode Island.

Zysman, J., & Kenney, M. (2018). The next phase in the digital revolution: Intelligent tools, platforms, growth, employment. Communications of the ACM, 61(2), 54–63. https://doi.org/10.1145/3173550

Apéndice

A1. Códigos implementados con *Python*

https://colab.research.google.com/drive/1PaXwEL09qubbsvu4agUMiLGhwER9MIY_?usp=sharing