INFORME TÉCNICO

GUÍA PARA EL RESGUARDO DE LOS DATOS PERSONALES EN EL DESARROLLO E IMPLEMENTACIÓN DE PLATAFORMAS DE DATOS ABIERTOS POR PARTE DE LOS ÓRGANOS DE LA ADMINISTRACIÓN DEL ESTADO

DJ/UNR/30.06.2020

I. DATOS ABIERTOS. DESCRIPCIÓN.

1. Antecedentes.

En la década de 1990, como resultado del desarrollo y expansión de las tecnologías digitales, surgen los conceptos de "gobierno 2.0" y "gobierno abierto", dando cuenta del creciente uso por parte de la Administración del Estado de plataformas virtuales y otras herramientas informáticas, con el objetivo de mejorar la eficiencia y eficacia del sector público y promover un mayor involucramiento de la ciudadanía en la gestión gubernamental.

Unido a la idea de gobierno abierto, surge el concepto de "datos abiertos", que busca permitir al público general el acceso y uso gratuito -en formatos electrónicos estandarizados- de la información generada o gestionada por el Estado (información del sector público), para cualquier tipo de finalidad. En este sentido, la *Open Data Charter* los define como aquellos datos digitales "que son puestos a disposición con las características técnicas y jurídicas necesarias para que puedan ser usados, reutilizados y redistribuidos libremente por cualquier persona, en cualquier momento y en cualquier lugar"[1]. Los principales elementos comprendidos en el concepto de datos abiertos son los siguientes:

- disponibilidad y oportunidad;
- empleo de formatos estandarizados, interoperables y comprensibles, que faciliten su reutilización[2] y redistribución; y,
- acceso universal, gratuito y sin propósitos acotados[3].

[1] https://opendatacharter.net/principles-es/

[2] La Directiva 2003/98/CE sobre la reutilización de la información del sector público, define "reutilización", en su artículo 2(4), como "el uso de documentos que obran en poder de organismos del sector público por personas físicas o jurídicas con fines comerciales o no comerciales distintos del propósito inicial que tenían esos documentos en la misión de servicio público para la que se produjeron. El intercambio de documentos entre organismos del sector público en el marco de sus actividades de servicio público no se considerará reutilización".

[3] La *Open Data Charter* contempla seis principios base para el acceso a los datos y para su publicación y uso, disponiendo que éstos deben ser: (i) abiertos por defecto; (ii) oportunos y exhaustivos; (iii) accesibles y utilizables; (iv) comparables e interoperables; (v) para mejorar la gobernanza y la participación ciudadana; y, (vi) para el desarrollo incluyente y la innovación. Por otra parte, puede sostenerse que los datos abiertos deben reunir las siguientes características: (i) disponibilidad en línea, lo que redunda en simplificar el acceso a los mismo, mediante su descarga; (ii) legibilidad por máquina, facilitando el análisis y la fusión de grandes conjuntos de datos heterogéneos, de manera automatizada; y, (iii) uso bajo una licencia abierta, la que impone restricciones mínimas al usuario.

En términos generales, es posible distinguir cuatro impulsores principales de la demanda por la apertura de datos públicos: (i) promover la transparencia y la rendición de cuentas; (ii) permitir la gobernanza participativa en la toma de decisiones; (iii) facilitar la innovación y el crecimiento económico; y, (iv) permitir nuevos usos por el propio sector público[4].

En cuanto a los fines que se pueden perseguir con el uso de estos datos, se pueden mencionar, a modo ejemplar, la posibilidad de exigir derechos o acceder a beneficios y servicios entregados por el Estado; ejercer un mayor control por parte de la sociedad respecto de las autoridades; facilitar el ejercicio de la actividad periodística; promover el desarrollo de investigaciones científico-académicas; y, contribuir a un mejor desempeño de las actividades comerciales o empresariales, generando valor a partir de dichos datos.

2. <u>Relevancia</u>.

La importancia de los datos abiertos radica en que diversos órganos públicos recopilan y almacenan una gran cantidad de datos, de diversa índole (personales, no personales y mixtos), con el objetivo de cumplir sus funciones y en ejercicio de sus competencias legales. En este sentido, es posible afirmar que el Estado es uno de los mayores procesadores de información relativa a las personas. En este sentido, los organismos públicos tendrían el deber de disponibilizar los datos que producen y almacenan, por cuanto éstos fueron recolectados o generados con los recursos aportados por los propios ciudadanos, en su calidad de contribuyentes[5].

Uno de los beneficios más evidentes del desarrollo de sistemas o plataformas de datos abiertos por parte del sector público es el aumento de los niveles de transparencia de la actividad de los gobiernos, fortaleciendo la democracia y el control de la sociedad respecto del ejercicio de la función pública. Así, por ejemplo, la transparencia de los datos sobre gasto fiscal y el presupuesto permite un debate público informado, facilitando un mayor control sobre las operaciones de los organismos públicos, al posibilitar un seguimiento y escrutinio más exhaustivo de la gestión de sus recursos.

La reutilización de los datos en poder del sector público constituye, también, un factor que puede desempeñar un papel relevante en la generación de nuevo conocimiento, innovación

Asimismo, en orden a reducir las barreras para el uso de Datos Abiertos, se debe favorecer el uso de formatos neutros como CSV y RDF sobre los propietarios como Excel (Simperl, O'Hara & Gomer, 2016: 5). Puede entenderse que cuando los datos se liberan en formatos electrónicos reutilizables bajo una licencia abierta para su reutilización, se cumple efectivamente con los principios de los datos abiertos, esto es, que están al servicio de los objetivos de la innovación y el desarrollo económico, además de los valores de la transparencia y la rendición de cuentas en el gobierno (Conroy & Scassa, 2015: 179).

[4] Janssen, 2012.

[5] Con todo, no puede soslayarse el hecho que la publicación de los datos públicos puede traer aparejados costos adicionales, relacionados principalmente con su procesamiento, organización, anonimización y combinación.

y oportunidades en la Economía Digital, impulsada en gran medida por el procesamiento y análisis de datos.

A este respecto, la Comisión Europea estima que el valor económico directo total de la información en poder del sector público dentro de la Unión Europea (UE), aumentará de una base de referencia de 52 mil millones de euros en 2018, a 194 mil millones de euros en 2030[6]. Influye en dicho valor el mejoramiento de las capacidades técnicas necesarias para facilitar el acceso a la información recopilada o generada por el Estado, haciendo que ésta pueda ser compartida de manera rápida y asequible. Uno de los mayores impactos a los cuales se hace constante referencia con la reutilización de datos abiertos, es el ahorro de costos que podría significar tanto para el sector público como para los privados, así como la creación de nuevos puestos de trabajo especializados en la utilización de datos[7].

A nivel de investigación periodística, las políticas de datos abiertos tienen un efecto positivo considerable. El Consejo para la Transparencia ha identificado casi 80 casos en que se detectaron irregularidades públicas por parte de medios de comunicación, en base al acceso a información pública. Esta línea de trabajo se ha fortalecido considerablemente en los últimos años, en vista al desarrollo del periodismo de datos[8]. A nivel local, el medio periodístico CIPER tiene una amplia trayectoria en la materia, mientras que en el plano internacional, destacan los casos de La Nación Data en Argentina, The Guardian en Inglaterra y The New York Times en Estados Unidos.

Siguiendo una línea similar, se observa un creciente protagonismo de las organizaciones de la sociedad civil, quienes, asumiendo un rol de control de la actuación estatal, utilizan los datos públicos para solicitar rendición de cuentas a las autoridades. Un ejemplo de esto es el Observatorio Fiscal, que en base a los datos de la Dirección de Compras y Contratación Pública, ha identificado espacios de opacidad en sectores como salud, gasto electoral y difusión de organismos públicos[9].

En cuanto a la investigación científica, las técnicas de análisis de *big data*, creadas a partir de la recopilación intensiva de grandes volúmenes de datos heterogéneos, están permitiendo el desarrollo de proyectos basados en modelos empíricos-analíticos y el diseño de respuestas más eficaces, que van más allá del muestreo aleatorio, para representar a casi todos los componentes de un grupo objeto de estudio. De esta forma, se entiende que el valor de los datos reside en su uso, y que cuantos más datos se puedan aplicar a un problema, mejor será el resultado de la intervención[10].

A dicho respecto, el estudio *The State of Open Data* contiene un levantamiento de la noción y comportamiento de los científicos con respecto a los datos abiertos. Entre los principales

[6] https://ec.europa.eu/digital-single-market/en/open-data
[7] https://www.mundofranquicia.com/actualidad/noticias/big-data-se-posiciona-gran-motor-crecimiento-economico-europa/
[8] https://estrategia.gobiernoenlinea.gov.co/623/w3-article-62470.html
[9] https://observatoriofiscal.cl/
[10] Simperl, O'Hara & Gomer, 2016: 4.

hallazgos, se destaca que un 78% de los encuestados considera que citar bases de datos tiene el mismo valor que citar un artículo académico. Asimismo, el 79% de los encuestados apoyan en general la existencia de un mandato nacional para poner a disposición del público las investigaciones primarias.

En el ámbito político y democrático, por su parte, se sostiene que el acceso a los datos públicos tiene la capacidad de asegurar la eficacia y los mejores resultados de los programas públicos al menor costo; elevar los niveles de legitimidad y cumplimiento de las decisiones gubernamentales; asegurar la equidad de acceso a la formulación de políticas públicas, disminuyendo las barreras de entrada a los procesos de decisión; y, en general, fomentar la participación ciudadana[11]. En este sentido, la idea de un gobierno abierto está ligada al ideal de transparencia de las decisiones y actividades de las entidades públicas, siendo considerada como una condición previa para el ejercicio efectivo de los derechos y libertades políticas, así como para asegurar la efectiva responsabilidad de las autoridades. Por consiguiente, para evaluar, debatir y sancionar el comportamiento del sector público se requiere necesariamente del acceso a información precisa respecto a sus distintas actividades[12].

La corriente de datos abiertos no sólo involucra la información que obra en poder del Estado, sino también aquella gestionada por organismos internacionales, organizaciones no gubernamentales y ciertas entidades privadas. Así, puede destacarse la creación de sitios web especiales donde se publica información en formatos abiertos, tales como el Portal Europeo de Datos (que recopila metadatos a partir de información del sector público disponible en portales de datos de acceso público de distintos países europeos)[13], los datos de libre acceso del Banco Mundial[14], la plataforma HDX de las Naciones Unidas[15], o la plataforma de búsqueda de bases de datos abiertos de Google[16].

3. Portales de datos abiertos en Chile.

A continuación, presentamos algunos ejemplos de portales digitales administrados por órganos estatales, que permiten el acceso y reutilización de la información del sector público:

(i) La Dirección de Compras y Contratación Pública (ChileCompra) posee una plataforma de datos abiertos, http://datosabiertos.chilecompra.cl, la cual disponibiliza información de las compras públicas realizadas a través del sitio web de Mercado

[11] https://www.cepal.org/ilpes/noticias/paginas/3/54303/Datos_Abiertos_Un_Nuevo_Desafio_Gobiernos.pdf
[12] Zuiderveen Borgesius et al. 2015: 2084.
[13] https://www.europeandataportal.eu/es/about/european-data-portal
[14] https://datos.bancomundial.org/
[15] https://data.humdata.org/
[16] https://toolbox.google.com/datasetsearch

Público, para facilitar su análisis, monitoreo y fiscalización, así como el desarrollo de nuevas aplicaciones e informes.

De acuerdo con lo informado en dicho portal, los datos están orientados a un modelo de colaboración interinstitucional, con el objeto de ofrecer una mayor cantidad de datos en formatos definidos, complementándolos con otros datos existentes. Así, dicha plataforma muestra dos tipos de datos, los que se encuentran procesados y depurados mensualmente, y los datos en duro que no se encuentran procesados.

(ii) **El Ministerio Secretaría General de la Presidencia** dispone de un portal de datos públicos, http://datos.gob.cl, que incluye un catálogo de fuentes de datos, publicados por diferentes instituciones públicas, bajo un esquema estandarizado, con el objeto de facilitar al público la búsqueda y posterior utilización de dichos datos. Muchos de estos datos ya se encuentran disponibles en diversos sitios electrónicos de otros servicios públicos, pero este portal los reúne en un solo sitio web, centralizando y facilitando su búsqueda.

El objetivo de la herramienta es que las personas tengan acceso a la información en base a la cual el Gobierno toma decisiones para el diseño y ejecución de políticas públicas, fortaleciendo la democracia y la rendición de cuentas por parte de las autoridades.

(iii) **La Dirección de Presupuestos del Ministerio de Hacienda** ha puesto a disposición del público la plataforma Presupuesto Abierto, http://presupuestoabierto.gob.cl, que permite conocer información detallada sobre los recursos asignados y su ejecución mensual del Gobierno Central. Además, entrega información detallada por sectores, regiones y principales receptores de recursos o proveedores del Estado.

Esta plataforma publica y visualiza los datos transaccionales de ejecución del gasto fiscal de la Presidencia de la República, el Congreso Nacional, el Poder Judicial, la Contraloría General de la República, el Ministerio Público, y cada uno de los ministerios y sus respectivos servicios, de forma simple e interactiva. El portal transparenta los datos de ejecución presupuestaria que se genera a partir de la Ley de Presupuestos para los años 2016, 2017, 2018 y 2019 (a la fecha), publicando las transacciones de devengo y sus datos asociados. Además, la plataforma se encuentra integrada con ChileCompra, para visualizar el detalle de las órdenes de compra asociadas a las transacciones de devengo, en caso de que hayan sido registradas.

(iv) **El Ministerio de Economía, Fomento y Turismo** dispone de un portal de datos abiertos denominado "Economía Abierta", https://www.economiaabierta.cl, cuyo objetivo es poner a disposición del público, información y datos estadísticos elaborados por dicho Ministerio y sus servicios relacionados, con el propósito de que puedan ser utilizados para generar cruces de información, análisis y diversas investigaciones, tanto de índole académica, como empresarial y ciudadana.

Algunos de los datos que se encuentran disponibles en éste sitio web son: Registro de Empresas y Sociedades (Subsecretaría de Economía y EMT); Censo 2017 (INE); Encuesta Suplementaria de Ingresos (INE); Encuesta de Uso del Tiempo (INE); Encuesta Nacional Industrial Anual (INE); Encuesta de Presupuestos Familiares (INE); Solicitudes de Registro de Patentes (INAPI); Solicitudes de Registro de Marcas (INAPI); Reclamos de Consumidores (SERNAC); Encuesta de Precios de Medicamentos y otras encuestas de precios (SERNAC); Listados de Renegociaciones, Nómina de Martilleros, Nómina de Veedores y Nómina de Liquidadores (SUPERIR)[17].

(v) La Ilustre Municipalidad de Puente Alto cuenta con un portal de datos abiertos, https://datosabiertos.mpuentealto.cl, que contiene diversos catálogos de información. De acuerdo a lo informado en dicho portal, su objetivo es permitir a los ciudadanos disponer, de manera pública y gratuita, de datos que les sean relevantes y confiables, para utilizarlos a modo de aprendizaje, o para crear nuevos servicios. Se entiende que la apertura de estos datos permite: tomar mejores decisiones (realizando cruce de datos, basándose en evidencia y optimizando recursos); fomentar la transparencia de la gestión pública (permitiendo a la ciudadanía, organizaciones de la sociedad civil y quien lo desee monitorear y fiscalizar el trabajo realizado por las instituciones, otorgándoles mayor legitimidad); fomentar la innovación (creación de nuevos emprendimientos); y, apoyar la interoperabilidad.

Resulta importante destacar que la plataforma en cuestión identifica las diferentes etapas que comprende su proceso de disponibilización de conjuntos de datos abiertos: sensibilización acerca de sus beneficios, selección de datos con valor público, no publicación de información que contenga datos personales, contar con plataformas con licencia abierta y contar con un plan de publicación, publicación y promoción.

(vi) El Consejo para la Transparencia (CPLT) con el objetivo de mejorar el acceso a la información pública en formatos que permitan su reutilización por parte de distintos actores de la sociedad, ha impulsado desde noviembre de 2010, diferentes proyectos dirigidos a la disponibilización de información del sector público, considerando para ello, técnicas que permiten su disposición en las llamadas tres, cuatro y cinco estrellas de los datos abiertos (https://5stardata.info/es/), a saber:

- Tres estrellas: usa formatos no propietarios (ejemplo: CSV en vez de Excel).
- Cuatro estrellas: usa URIs para denotar cosas.
- Cinco estrellas: enlaza datos, para proveer contexto.

a) InfoLobby e InfoProbidad. En el marco de la Ley N°20.730 que regula el lobby, el CPLT tiene el rol de recibir, consolidar y publicar cierta información, poniendo a disposición de la ciudadanía la plataforma electrónica www.infolobby.cl, disponible desde noviembre de 2014, con los registros de agenda pública de las autoridades y funcionarios que son sujetos obligados de la ley. A mayo de 2020, cuenta con 436 mil

[17] https://digital.gob.cl/noticias/ministerio-de-economia-lanza-plataforma-de-datos-abiertos.

audiencias, 425 mil viales, 37 mil donativos, 18 mil lobistas y más de 300 mil autoridades funcionarios públicos, informados por 696 órganos públicos.

Respecto a InfoProbidad, la Ley de Probidad en la Función Pública y Prevención de los Conflictos de Intereses, Ley N°20.880, pone a disposición de la ciudadanía la declaración de patrimonio e intereses de quienes ocupan altos cargos en la función pública. Junto a lo anterior, la ley mandata a la Contraloría General de la República y al CPLT a disponer de las declaraciones de patrimonio e intereses de los funcionarios públicos obligados, en un portal accesible a toda la ciudadanía, en formato de datos abiertos y reutilizables. De esta forma, el año 2016, se crea el portal InfoProbidad, administrado por el CPLT, y que, a mayo de 2020, cuenta con 56 mil declaraciones de intereses y patrimonio informadas por 660 organismos públicos.

Tanto InfoLobby como InfoProbidad están construidos desde su diseño para el uso de datos abiertos, permitiendo la extracción de información en archivos de datos (3 estrellas de los datos abiertos), y en modalidad de web semántica o datos abiertos enlazados (5 estrellas de los datos abiertos). De esta forma, los datos de ambos portales pueden ser vinculados en línea, a otras fuentes de información. Cabe destacar que el 100% de los sitios operan sobre una base de datos semántica que puede ser reutilizada por cualquiera. Es decir, las mismas bases de datos que utilizan los sitios web de InfoLobby e InfoProbidad, son de acceso público y puede ser enlazada desde cualquier parte del mundo, utilizando el lenguaje de consulta de datos Sparql[18].

b) Portal de Transparencia. Mediante el Portal de Transparencia del Estado, publicado en mayo de 2013 y administrado por el CPLT, los organismos públicos gestionan las solicitudes de acceso a la información realizadas por los ciudadanos, y publican sus deberes de Transparencia Activa (TA), de acuerdo a lo estipulado en la Ley de Transparencia de la Función Pública y de Acceso a la Información de la Administración del Estado, aprobada por el artículo primero de la Ley N° 20.285 (Ley de Transparencia). A partir de esto, se publican en formato de tres estrellas de datos abiertos, información relativa a las solicitudes y las publicaciones en TA. Estos datos pueden ser accedidos en: https://www.portaltransparencia.cl/PortalPdT/web/guest/opendata

A mayo de 2020, el Portal de Transparencia cuenta con 833 organismos públicos incorporados a la plataforma, más de 900 mil solicitudes de acceso a la información, 142 mil ciudadanos registrados y 14 mil funcionarios públicos.

[18] Los sitios de InfoLobby e InfoProbidad, generaron el año 2018 un impacto positivo en el foro OCDE. Ambos fueron destacados –particularmente Infolobby– en la labor de promoción de la integridad en los países miembros. Como muestra de ello, el "Open Government Data Report – Enhancing Policy Maturity for Sustainable Impact" de la OCDE destacó el caso chileno. Ver: http://www.oecd.org/gov/open-government-data-report-9789264305847-en.htm, página 197.

II. MARCO NORMATIVO GENERAL APLICABLE A LA PROTECCIÓN DE LOS DATOS PERSONALES.

1. Concepto.

El derecho fundamental a la protección de los datos personales se encuentra reconocido como tal en el numeral 4° del artículo 19 de la Constitución Política de la República. El contenido de este derecho consiste, en términos simples, en la facultad que tiene cada individuo de controlar el procesamiento y flujo de aquella información que le concierne. Este derecho se concretiza, en el rango legal, en el artículo 4° de la Ley N°19.628, sobre Protección de la Vida Privada (LPVP), donde se establece que el tratamiento de los datos personales sólo puede efectuarse cuando el titular consiente expresamente en ello (por escrito) o la ley lo autorice.

En el ordenamiento chileno, la LPVP constituye el cuerpo normativo que regula específicamente el procesamiento de datos personales, el funcionamiento de las bases de datos, los derechos y deberes de los involucrados y un mecanismo de solución de controversias[19].

Las disposiciones de la LPVP son aplicables al tratamiento de los datos de carácter personal en registros o bancos de datos por organismos públicos o por particulares[20]. El concepto de "tratamiento de datos personales" está definido, de manera amplia, en el artículo 2°, letra o), del referido cuerpo legal, siendo entendido como "cualquier operación o complejo de operaciones o procedimientos técnicos, de carácter automatizado o no, que permitan recolectar, almacenar, grabar, organizar, elaborar, seleccionar, extraer, confrontar, interconectar, disociar, comunicar, ceder, transferir, transmitir o cancelar datos de carácter personal, o utilizarlos en cualquier otra forma.".

Respecto a la definición de "dato personal" contenida en nuestro ordenamiento jurídico, también es amplia. Según establece el literal f) del artículo 2° de la LPVP, los datos de carácter personal son aquellos "relativos a cualquier información concerniente a personas naturales, identificadas o identificables.". Así, a modo ejemplar, cabe dentro de este concepto el nombre y apellidos de una persona, su domicilio, su dirección de correo electrónico, su número de documento nacional de identidad, sus antecedentes comerciales y financieros, datos de localización, o la identificación de la dirección de protocolo de internet (IP) desde la cual navega en la web. Respecto del formato, el concepto de datos

[19] Viollier, 2017: 7.

[20] Según establece el literal m) del artículo 2° de la LPVP, registro o banco de datos es "el conjunto organizado de datos de carácter personal, sea automatizado o no y cualquiera sea la forma o modalidad de su creación u organización, que permita relacionar los datos entre sí, así como realizar todo tipo de tratamiento de datos.". Por ejemplo, se pueden mencionar: los registros de personal de una institución, los de datos de usuarios, los de datos de beneficiarios de subsidios, los de proveedores, un documento o tabla excel en el que se incluya distintos nombres y direcciones de participantes de un evento, un conjunto de currículos en formato digital insertos en una carpeta catalogados por nombre, etc.

personales abarca información alfabética, numérica, gráfica, fotográfica o sonora, por citar algunas, pudiendo estar contenida en cualquier soporte, tanto físico como digital.

Por otra parte, debe entenderse que una persona es identificable cuando su identidad pueda determinarse directa o indirectamente mediante un identificador (por ejemplo, un nombre, un número de registro, datos de localización o un identificador en línea) o mediante el uso de uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de las personas. En este sentido, se habla de identificadores directos (características dentro de un conjunto de datos que, por sí mismos, identifican a los individuos) e identificadores indirectos o cuasi-identificadores (características dentro de un conjunto de datos que, en combinación con otros datos, identifican a los individuos). La capacidad de vincular características a través de los conjuntos de datos y aprender sobre los individuos se conoce como el "efecto mosaico"[21].

Existen también ciertas categorías especiales de datos, sujetos a mayores niveles de resguardo, denominados "datos sensibles", que son definidos, en el literal g) del artículo 2º de la LPVP, como "aquellos datos personales que se refieren a las características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidad, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual.".

Al tratarse de una definición legal de carácter abierto, el concepto de dato personal sensible puede abarcar aspectos tan disímiles como la información médica de las personas, sus hábitos de conducta o su vida afectiva. De esta manera, al momento de calificar un dato personal como sensible, los organismos públicos deben tener presente, al menos, las siguientes categorías: (i) datos que se refieren a características físicas de una persona, tales como datos biométricos, muestras y datos biológicos, datos de salud (sea física o psíquica), o datos sobre estados de ánimo, entre otros; (ii) datos que se refieren a características morales de una persona, tales como información sobre orientación o preferencia sexual, creencias o convicciones religiosas, éticas o políticas, entre otros de similar naturaleza; y, (iii) datos que se refieren a hechos o circunstancias de su vida privada o intimidad, tales como los hábitos personales, la información sobre desplazamiento y geolocalización, o sus redes de amistad y contacto, entre otros.

Respecto de estos datos, el artículo 10 de la LPVP prohíbe su tratamiento o comunicación, salvo cuando la ley lo autorice, exista consentimiento del titular o sean datos necesarios para la determinación u otorgamiento de beneficios de salud que correspondan a sus titulares.

[21] Green et al., 2017: 20.

2. <u>Tratamiento de datos personales por organismos públicos</u>.

Según dispone el artículo 2°, letra n), de la LPVP, los organismos públicos pueden revestir la calidad de responsables de una base de datos de carácter personal, cuando les compete las decisiones relacionadas con el tratamiento de la misma.

Luego, el artículo 20 de la LPVP contiene una habilitación o autorización a los organismos públicos para tratar datos personales respecto de las materias de su competencia, sin requerir el consentimiento del titular. Con todo, dicha norma dispone expresamente que dichos tratamientos deben sujetarse a las reglas contenidas en la misma ley, cobrando especial relevancia las siguientes disposiciones:

a) **Principio de finalidad:** según dispone el inciso primero del artículo 9° de la LPVP, las operaciones de tratamiento que se realicen respecto de datos personales deberán circunscribirse estrictamente a los fines para los cuales hubieran sido inicialmente recolectados. La referida finalidad en el caso de órganos de la Administración del Estado estará determinada en función de las materias propias de su competencia.

b) **Principio de calidad de los datos:** los datos tratados deben ser exactos, actualizados y responder con veracidad a la situación real de su titular. El organismo o servicio público responsable de la base de datos debe, sin necesidad de requerimiento del titular de los mismos: eliminar los datos caducos y aquellos que estén fuera de su competencia; bloquear los datos personales cuya exactitud no pueda ser establecida o cuya vigencia sea dudosa y respecto de los cuáles no corresponda su cancelación; y, modificar los datos inexactos, equívocos o incompletos.

c) **Principio de proporcionalidad:** implica que sólo pueden recabarse aquellos datos que sean necesarios para conseguir los fines que justifican su recolección. Se entiende que se cumple con el principio de proporcionalidad cuando: el o los datos que se recolecten, así como su posterior tratamiento, sean adecuados o apropiados a la finalidad que lo motiva; sean pertinentes o conducentes para conseguir la referida finalidad; y, no excesivos, en el sentido que no exista otra medida más moderada para la consecución de tal propósito con igual eficacia. En aplicación de este principio, los órganos o servicios públicos deben optar, de entre los diversos tratamientos que le permitan conseguir los fines pretendidos dentro del ámbito de sus competencias, por aquel que menor incidencia tenga en el derecho a la protección de datos personales y por la utilización de los medios menos invasivos.

d) **Principio de seguridad y responsabilidad:** conforme a lo establecido en el artículo 11 de la LPVP, el responsable de la base de datos deberá cuidar de ellos con la debida diligencia, haciéndose responsable de los daños[22]. De esta forma, los órganos de la Administración del Estado debiesen aplicar en sus actividades de tratamiento de datos

[22] Sin embargo, cabe advertir que "la ley no establece estándares de cuidado o medidas concretas que dichos responsables deban tomar a fin de velar por la seguridad de los datos o prevenir su daño" (Viollier, 2017: 23).

personales medidas de seguridad, técnicas y organizativas que garanticen la confidencialidad, integridad y disponibilidad de la información. Para ello, es necesario establecer diversos niveles de seguridad, atendiendo al tipo de dato almacenado (a título ejemplar, respecto de los datos sensibles deberán adoptarse niveles de seguridad más altos que en relación a aquellos que no poseen dicha calidad).

e) **Principio de confidencialidad:** según prescribe el artículo 7° de la LPVP, las personas que trabajan en el tratamiento de datos personales o tengan acceso a éstos de otra forma, están obligadas a guardar secreto sobre los mismos, cuando provengan o hayan sido recolectados de fuentes no accesibles al público, como asimismo sobre los demás datos y antecedentes relacionados con el banco de datos, obligación que no cesa por haber terminado sus actividades en ese campo.

f) **Principio de información:** de acuerdo a lo dispuesto en los artículos 3°, 4° y 20 de la LPVP, y aunque los organismos públicos estén facultados para efectuar tratamientos de datos de carácter personal sin consentimiento del titular de los mismos respecto de materias de su competencia, estos debiesen, previamente a su recolección, informar a su titular acerca del órgano responsable de la base de datos, de la finalidad perseguida con el tratamiento de la información, de la posible comunicación a terceros y de los derechos que pueden ser ejercidos por ellos. Lo anterior, poniendo en conocimiento de los respectivos titulares las políticas de protección de datos personales que se adoptarán en los tratamientos respectivos.

Asimismo, los órganos de la Administración del Estado tienen el deber de garantizar a los titulares de datos personales el ejercicio de los derechos establecidos en el artículo 12 de la LPVP –esto es, derechos de acceso, rectificación o modificación, cancelación o eliminación y bloqueo de datos- teniendo presente las características de independencia, gratuidad y sencillez.

Cabe advertir, también, la obligación especial de los órganos de la Administración del Estado de inscribir todos los bancos de datos personales que obren en su poder en el Registro de los Bancos de Datos Personales a cargo de Organismos Públicos que lleva el Servicio de Registro Civil e Identificación, de acuerdo a lo establecido en el artículo 22 de la LPVP[23].

Para facilitar el cumplimiento de las obligaciones que la LPVP contempla respecto de los órganos de la Administración del Estado, el Consejo para la Transparencia sugiere –en sus Recomendaciones sobre Protección de Datos Personales por parte de los órganos de la Administración del Estado, publicadas en el Diario Oficial con fecha 14 de septiembre de 2011- que las distintas autoridades, jefaturas o jefes superiores de los órganos o servicios de la Administración del Estado, designen a un funcionario o funcionaria de dicha repartición

[23] Norma complementada por el Decreto Supremo N° 779, de 2000, del Ministerio de Justicia, que aprobó el Reglamento del Registro de Bancos de Datos Personales a cargo de Organismos Públicos, y por la Resolución (E) N° 1540, de 2010, del Servicio de Registro Civil e Identificación.

para desempeñarse como encargado o encargada de protección de datos y constituya un contacto efectivo en la materia con el CPLT.

III. REGULACIÓN DE LOS DATOS ABIERTOS.

En términos generales, los deberes de los organismos públicos relativos a la divulgación de información como datos abiertos, y sus respectivas reglas, tienden a no estar codificadas en normas de carácter legal. Más bien, las políticas de datos abiertos son usualmente promovidas mediante normas de rango administrativo, las que vienen a definir sus objetivos, metas e instrucciones[24].

No obstante lo anterior, muchas iniciativas de datos abiertos encuentran sus fundamentos en los principios que regulan el derecho de acceso a la información pública, la que dan cuenta de obligaciones de divulgación más genéricas, que no suelen establecer la forma cómo debe ponerse a disposición del público la información sobre la cual recae este derecho de acceso (por ejemplo, en un formato considerado abierto, legible por máquina)[25].

1. Datos abiertos en el derecho comparado.

En materia de disponibilización de información del sector público en formatos abiertos, destaca el caso de la Unión Europea y su Directiva 2003/98/CE sobre la reutilización de la información del sector público (revisada por la Directiva 2013/37/UE).

Cabe hacer presente que el marco europeo sobre datos abiertos se diferencia claramente de su normativa sobre la libertad de información, al abordar la apertura y transparencia de la información del sector público no como un fin en sí mismo, sino más bien, como un medio que puede "desempeñar una función importante a la hora de impulsar el desarrollo de nuevos servicios basados en formas novedosas de combinar y utilizar esa información, estimular el crecimiento económico y promover el compromiso social"[26]. Por lo tanto, se señala que el eje de esta legislación sobre la reutilización de la información del sector público es de carácter económico, reforzado por el objetivo de armonizar la forma en que los datos públicos se hacen reutilizables, de modo que se pueda acceder a ellos, combinarlos y vincularlos fácilmente[27].

El principio fundamental contenido en este marco normativo es que los Estados miembros de la UE garanticen en aquellos casos en que se permita la reutilización de la información del sector público, que la apertura comprenda fines comerciales y no comerciales, con restricciones mínimas o nulas de carácter jurídico, técnico o económico, y armonizando las condiciones básicas para su acceso y uso, con el fin de garantizar que dichas condiciones sean equitativas, proporcionadas y no discriminatorias. Para ello, se establecen una serie de

[24] Zuiderveen Borgesius et al. 2015: 2093.
[25] *Ibid.* 2098.
[26] Directiva 2013/37/UE.
[27] Wiebe & Dietrich, 2017: 215.

criterios rectores, tanto sustantivos (v.g. los tipos de documentos que pueden estar sujetos a reutilización, principio de no discriminación), como formales (v.g. solicitudes de reutilización, tipos de licencias y formatos que facilitan la disponibilización de la información en entornos digitalizados, reglas sobre tarificación). En su mayoría, estos criterios son procedimentales, dejando en gran medida en manos de la legislación doméstica de los Estados miembros la determinación de los documentos que, en definitiva, cumplen con los requisitos para ser accesibles al público[28].

Este marco normativo reconoce que su implementación y aplicación deber darse en pleno cumplimiento de los principios relativos a la protección de los datos de carácter personal, estableciendo explícitamente que no modifica las obligaciones y los derechos contenidos en la normativa de la UE sobre protección de datos personales[29]. Se deben tener presente, también, las excepciones en materia de protección de datos que la propia Directiva establece, tales como los documentos cuyo acceso está excluido según la normativa de protección de datos; documentos cuyo acceso está restringido según la normativa de protección de datos; y partes de documentos que son accesibles conforme a la normativa de protección de datos, y cuya reutilización, según lo establecido por la ley, es incompatible con la protección de los individuos en lo relativo al tratamiento de los datos de carácter personal[30].

A dicho respecto, resulta relevante tener presente la Opinión 06/2013 sobre datos abiertos y reutilización de información del sector público, del Grupo de Trabajo del Artículo 29 (GT29)[31]. El objetivo de esta opinión es ayudar a garantizar un entendimiento común del

[28] *Ibid.* 211.

[29] Jaatinen, 2016: 3.

[30] Cabe hacer presente que la Directiva 2003/98/CE, modificada por la Directiva 2013/37/UE, queda derogada con efectos a partir del 17 de julio de 2021, por la Directiva (UE) 2019/1024, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público, que tiene por finalidad afrontar los obstáculos derivados de la amplia reutilización de la información del sector público en la UE y actualizar el marco legislativo con los avances en las tecnologías digitales, estimulando la innovación, en especial en lo que respecta a inteligencia artificial. Los cambios introducidos por esta nueva normativa están enfocados, entre otros elementos, en la disponibilización de los denominados "datos dinámicos" (documentos en formato digital, sujetos a actualizaciones frecuentes o en tiempo real, debido, en particular, a su volatilidad o rápida obsolescencia) para su reutilización inmediatamente después de su recopilación; y, en aumentar el acceso a conjuntos de datos públicos catalogados como de "alto valor" (documentos cuya reutilización está asociada a considerables beneficios para la sociedad, el medio ambiente y la economía, en vista a su idoneidad para la creación de servicios de valor añadido, aplicaciones y puestos de trabajo nuevos, dignos y de calidad, y del número de beneficiarios potenciales de los servicios de valor añadido y aplicaciones basados en tales conjuntos de datos).

[31] Órgano de carácter consultivo e independiente, creado por la Directiva 95/46/CE relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, e integrado por las autoridades de protección de datos de todos los Estados miembros de la UE, el Supervisor Europeo de Protección de Datos y la Comisión Europea. Sus funciones incluían, entre otras, estudiar toda cuestión relativa a la aplicación de las disposiciones nacionales tomadas para la aplicación de la Directiva 95/46/CE con vistas a contribuir a su aplicación homogénea; emitir un dictamen destinado a la Comisión Europea sobre el nivel de

marco jurídico aplicable y ofrecer una orientación coherente y ejemplos de buenas prácticas relativas a la aplicación de la normativa de la UE sobre datos abiertos, en lo que respecta al tratamiento de datos personales. Así, la Opinión 06/2013 pone énfasis en el hecho que no todos los datos personales "públicamente disponibles" debiesen ser publicados como datos abiertos, por cuanto la obligación de reutilización, según la Directiva 2003/98/CE, se entiende sin perjuicio de los requisitos de protección de datos. De esta forma, el principio de reutilización no es de aplicación automática cuando está en juego el derecho a la protección de datos personales.

El GT29 se refiere a la importancia de una evaluación de impacto de protección de datos antes de proceder con la apertura datos del sector público para su reutilización, haciendo hincapié en la necesidad de adherir a los principios de privacidad por diseño y por defecto, garantizando la debida protección de los datos en una etapa temprana, y teniendo especialmente presentes los principios de finalidad, proporcionalidad y minimización en el tratamiento de datos personales.

Para el GT29 es indispensable considerar la normativa sobre protección de datos como una limitación en la selección de los datos que pueden o no ser objeto de acceso[32] y reutilización, por cuanto no suelen permitir que los organismos del sector público revelen públicamente los datos personales reunidos para un fin distinto de aquel que motivó su recopilación, incluyendo su posible reutilización en el marco de iniciativas de datos abiertos. En lugar de datos de carácter personal, suelen publicarse datos estadísticos derivados de éstos, reduciendo al mínimo los riesgos de divulgación involuntaria de datos personales. Así, por ejemplo, se hace referencia a la reutilización de conjuntos de datos agregados y anonimizados derivados de datos personales, los que no deberían permitir la reidentificación de las personas. En definitiva, la legislación de protección de datos tiene un importante papel que desempeñar en la determinación del umbral bajo el cual es seguro liberar datos anonimizados y agregados como parte de una iniciativa de datos abiertos, empleando técnicas adecuadas.

2. Datos Abiertos en la normativa chilena.

El año 2012, el Gobierno de Chile fijó como un eje fundamental en su gestión, el impulso de una agenda de modernización de las instituciones públicas, haciendo que la transparencia, la participación y la colaboración fueran accesibles por la ciudadanía.

protección existente dentro de la Comunidad Europea y en países terceros; y, asesorar a la Comisión Europea sobre cualquier proyecto de modificación de la referida directiva. Con la entrada en vigencia y aplicación del Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE, el GT29 fue reemplazado por el Comité Europeo de Protección de Datos.

[32] Con todo, el GT29 reconoce que la legislación sobre la libertad de información puede exigir la divulgación de datos personales, y que el interés por la transparencia en algunas situaciones puede prevalecer sobre las preocupaciones relativas a la protección de los datos y la privacidad.

En este contexto, destaca el Instructivo Presidencial N°005 de 2012, que imparte instrucciones sobre Gobierno Abierto, abordando la publicación de datos por parte de la Administración del Estado. El objetivo de este Instructivo es consolidar un Portal de Gobierno Abierto y una Política de Datos Abiertos, promoviendo el acceso expedito, abierto y sin restricciones de uso a conjuntos de datos gubernamentales, permitiendo su reutilización por parte de terceros. Así, se dispuso que el Ministerio Secretaría General de la Presidencia, a través de su Unidad de Modernización y Gobierno Digital[33], gestionara un catálogo de Datos Abiertos del Gobierno de Chile.

En lo que respecta a los estándares de seguridad técnicos y organizativos de dicho catálogo, éstos se encuentran regulados en la norma técnica aprobada por el Ministerio Secretaría General de la Presidencia, de febrero de 2013. La mencionada norma técnica contempla expresamente, en su punto 4.2, la obligación de dar protección a los datos de carácter personal, de conformidad con lo dispuesto en la LPVP, disponiendo sobre el particular que "se debe tener especial consideración en la liberación de datos gubernamentales que contienen datos personales y/o datos sensibles de los ciudadanos, de forma de protegerlos.".

Adicionalmente, se debe tener presente que la LPVP da cuenta de una regla que tiene impacto en materia de disponibilización de datos en formatos abiertos. A dicho respecto, su artículo 3° establece que en la comunicación de los resultados de encuestas, estudios de mercado o sondeos de opinión pública u otros instrumentos semejantes, respecto de las cuales se hayan recolectado datos personales, se debe omitir las señas que puedan permitir la identificación de las personas consultadas.

IV. DATOS ABIERTOS Y PROTECCIÓN DE DATOS PERSONALES.

En general, se afirma que las iniciativas de datos abiertos deben apreciarse en estrecha relación con los conceptos de privacidad, protección de datos personales y transparencia[34]. No obstante lo anterior, en una primera etapa, la discusión sobre datos abiertos e información del sector público se concentró en elementos relacionados con la disponibilidad de la información, las condiciones de acceso y la reusabilidad de la información, más que en los principios legales y normas que restringen estos flujos de información, tales como aquellas vinculadas a la protección de los datos personales[35].

Si bien inicialmente se consideraba que los datos abiertos solo se trataban de datos no personales y, por lo tanto, no planteaban problemas en relación a la protección de la vida privada y de los datos personales, la posibilidad de reidentificar a las personas a partir de

[33] La Ley N°21.050, publicada en el Diario Oficial con fecha 7 de diciembre de 2017, modificó la Ley N°18.993, que crea el Ministerio Secretaría General de la Presidencia, incorporando en su estructura organizativa la División de Gobierno Digital, que reemplazo a la referida Unidad de Modernización y Gobierno Digital (artículo 3°, literal c).
[34] Jaatinen, 2016: 2.
[35] Pagallo & Bassi, 2013: 179.

conjuntos de datos no identificados -utilizando datos provenientes de múltiples fuentes, que facilitan el efecto mosaico- agudizó las preocupaciones sobre la privacidad, protección de datos personales y los datos abiertos[36]. Este problema se intensifica respecto de los datos abiertos granulares, los que, junto con posibilitar a sus usuarios análisis variados y detallados, a menudo incluyen información de carácter personal[37].

Los responsables políticos y los actores de la sociedad civil reconocen el impacto que los datos abiertos tienen en materia de privacidad y protección de datos personales[38]. Muchos de los datos que han sido liberados, o están siendo considerados para su liberación en formatos abiertos dicen relación con el comportamiento y las características de ciudadanos individuales, dejando en evidencia las tensiones entre los datos abiertos y la privacidad[39]. Así, por ejemplo, la *Open Data Charter* reconoce específicamente, en el desarrollo de su primer principio, la necesidad de observar las leyes locales y normas reconocidas internacionalmente sobre privacidad y confidencialidad. Asimismo, el *G8 Open Data Charter* dispone en esta materia la necesidad de observar la legislación nacional e internacional, en particular en lo que respecta a la propiedad intelectual, la información de identificación personal y la información sensible.

1. Datos públicos y datos no públicos.

Si bien el hecho que un dato se encuentre en posesión de un organismo estatal constituye una condición necesaria para que éste pueda ser considerado un dato abierto, se debe tener presente que no toda la información procesada por el Estado puede ser de acceso libre y reutilizable, debiendo distinguirse claramente aquella información que no reviste el carácter de "pública".

Los organismos públicos almacenan diversas categorías de información, incluyendo tanto datos que revisten la calidad de públicos como datos confidenciales o reservados. Así, por ejemplo, existen datos sujetos a causales de secreto o reserva que impiden su divulgación. De igual forma, los datos de carácter personal no revisten el carácter de "abiertos" o de ser "reutilizables" por cualquier persona y para cualquier propósito, salvo limitadas excepciones (por ejemplo, cuando provienen o se recolectan de fuentes accesibles al público).

2. Diagnóstico de posibles riesgos en materia de protección de datos personales a consecuencia de la disponibilización de datos abiertos.

En el desarrollo de iniciativas inspiradas en la idea de datos abiertos, es posible identificar diversas complejidades que pueden incidir en el efectivo resguardo de los datos personales:

[36] Scassa, 2018: 4.
[37] Green et al., 2017: 4.
[38] Zuiderveen Borgesius et al. 2015: 2107.
[39] Wood et al., 2016: 3.

(i) El concepto de dato personal es amplio y dinámico, lo que conlleva que estos tipos de datos pueden estar mezclados con información no personal, en bancos mixtos[40]. Esto dificulta la segregación de la información, especialmente en aquellos casos en que las plataformas de datos abiertos deben depurar o filtrar directamente las bases de datos mixtas, antes de su divulgación al público.

(ii) Las políticas de datos abiertos fomentan la reutilización de datos para fines no previstos con antelación, al momento de su recopilación. Así, el eventual uso secundario de información de carácter personal conlleva riesgos para la privacidad y la protección de datos personales, pudiendo vulnerar el principio de finalidad, una de las bases de los marcos normativos sobre protección de datos personales. Del principio de especificación de finalidad se desprende que los datos personales sólo deben recogerse para un fin especificado de antemano, y que esos datos no deben utilizarse para fines incompatibles[41]. Esto, se contrapone al ideal de los datos abiertos, esto es, que pueden ser usados, modificados y compartidos libremente por cualquiera para cualquier propósito, lo que, incluso, comprende propósitos comerciales.
Por otra parte, la reutilización comercial de datos abiertos, particularmente en los entornos digitales y de *big data*, puede tener un impacto negativo en la privacidad[42]. Las nuevas herramientas de *data mining* (importante especialmente para el análisis de "metadatos"[43]) y de cruce de distintas bases de datos han facilitado enormemente la re-asociación de información que inicialmente era presentada de manera disociada[44]. Asimismo, se verifica una importante disminución de los costos de recopilación, almacenamiento, procesamiento, análisis y difusión de grandes cantidades de datos[45].

(iii) Información que, aparentemente, no reviste la calidad de dato personal puede, de manera indirecta, estar vinculada o entregar algún indicio acerca de una persona individualizable. Es posible que información sobre personas específicas pueda generarse por inferencia, a partir de datos considerados inicialmente como no

[40] Así, por ejemplo, a veces es difícil separar la información concerniente a una persona jurídica respecto de información sobre personales naturales. Por otra parte, incluso los datos supuestamente no personales pueden proporcionar información sobre un individuo (Zuiderveen Borgesius et al. 2015: 2121).
[41] Zuiderveen Borgesius et al. 2015: 2109.
[42] Scassa, 2018: 8.
[43] Los metadatos son información estructurada que describe, explica, localiza o facilita de alguna manera la recuperación, el uso o la gestión de un recurso de información. Los metadatos suelen denominarse datos sobre datos o información sobre información. En una base de datos de correos electrónicos, por ejemplo, los metadatos contienen el remitente, el destinatario y la marca de tiempo de los correos electrónicos. Si bien los metadatos de los correos electrónicos no contienen el contenido de los mismos, pueden revelar patrones sobre la forma en que las personas actúan. Así, los metadatos a menudo comprenden registros de comportamiento (Green et al., 2017: 21).
[44] Al mismo tiempo que los editores de datos abiertos luchan por identificar y abordar posibles problemas de privacidad, los gobiernos publican habitualmente un gran volumen de información, a menudo de carácter personal, sobre la base de políticas elaboradas antes de la era del *big data* y, en algunos casos, incluso antes del surgimiento de Internet (Scassa, 2018: 10).
[45] Wood et al., 2016: 4.

personales, contenidos en un conjunto de datos abiertos[46]. La disponibilidad de todos estos datos contribuye a las cuestiones de la identificabilidad de las personas, debido a la posibilidad de combinar diferentes fuentes de datos para llevar a cabo procesos de reidentificación[47]. Adicionalmente, se debe tener presente que las bases de datos pueden ser manipuladas y combinadas de maneras complejas e impredecibles[48].
El uso combinado de información proveniente de diversas fuentes de datos abiertos, sumado al uso de técnicas de análisis de grandes volúmenes de datos y herramientas de aprendizaje automático plantea serios riesgos a la protección de datos personales. Así, a medida que el análisis de datos se hace más sofisticado, y que el volumen de otros datos disponibles crece exponencialmente, los riesgos de reidentificación en conjuntos de datos anonimizados pueden ser extremadamente altos[49]. En este sentido, se debe tener presente que muchas compañías y otros grupos recopilan información para construir perfiles digitales con las características, hábitos y preferencias de una persona determinada, de manera tal que los datos abiertos pueden aportar nueva información a estos perfiles[50].

(iv) En muchas ocasiones, los titulares de datos no tienen la opción de no revelar o no entregar ciertos datos a los organismos públicos, por constituir un requisito para acceder a un cierto beneficio o servicio, o para ejercer un derecho, o por objeto de una intervención sancionadora del Estado. Asimismo, ciertas clases de datos personales son generadas por un determinado organismo público (v.g. documentos de identificación, permisos o licencias habilitantes para desarrollar una actividad), lo que dificulta la posibilidad que el titular de datos pueda ejercer de manera efectiva un control o autodeterminación sobre el mismo.

(v) Existe un déficit de medidas de seguridad adecuadas. En materia de seguridad informática, en términos generales, las medidas existentes estarían fragmentadas. Asimismo, es posible apreciar una carencia de resguardos apropiados en las plataformas digitales, como también en las soluciones técnicas y estructura organizativa de las instituciones frente a posibles ataques, incidentes o vulnarabilidades. Por otra parte, existirían problemas de configuración tecnológica, que aumentan los riesgos informáticos. Estos aspectos deficitarios se darían tanto en las operaciones de publicación de datos abiertos, en particular, como en el tratamiento

[46] *Ibid.* 7. Esta posibilidad dependerá de una serie de factores, por ejemplo, si existen "datos auxiliares" disponibles, si hay controles de acceso o de consulta sobre ellos, y si hay cortafuegos entre el conjunto de datos y los datos auxiliares. Asimismo, aunque una base de datos abiertos no incluya datos personales en el momento de su publicación, podría dar cuenta de datos relativos a personas identificables o determinables a medida que aumenta la cantidad de datos auxiliares. En este sentido, correspondería establecer si es posible aislar los registros sobre una persona de la base de datos; si es posible vincular los registros sobre la misma persona en la respectiva base de datos o entre bases de datos; y, si es posible deducir, con una probabilidad suficientemente alta, el valor de un atributo de una persona (Simperl, O'Hara & Gomer, 2016: 11).
[47] Zuiderveen Borgesius et al. 2015: 2121.
[48] Green et al., 2017: 3.
[49] Scassa, 2018: 5.
[50] Green et al., 2017: 18.

de bases de datos personales, en general, en distintos ejes: políticas y protocolos existentes; capacidades instaladas en las instituciones para enfrentar los problemas de seguridad informática y gestionar adecuadamente herramientas tecnológicas implementadas.

El principio de seguridad requiere la implementación de medidas de seguridad adecuadas para salvaguardar los datos personales, evitando la divulgación y el acceso no autorizados, u otro uso no permitido. Se debe hacer hincapié en el hecho que una vez que un dato es revelado o puesto a disposición del público general para su reutilización, no existe la posibilidad de recuperar el control sobre dicho dato[51]. Actualmente, la posibilidad de realizar copias y recircular la información es prácticamente ilimitada, y las consecuencias potenciales de esta pérdida de control se magnifican con las técnicas de *big data*[52].

3. Desafío.

El acceso y reutilización de datos abiertos debe darse en el marco de la protección de datos personales, cuyos derechos y obligaciones deben estar en el centro del diseño e implementación de estas iniciativas. Los estándares comparados exigen que el acceso a la información pública, la protección de la vida privada y de los datos personales sean conciliados adecuadamente, lo que requiere del sector público el empleo de herramientas que permitan facilitar la tarea de disponibilizar la información sin afectar su deber de resguardar los datos personales contenidos en las bases de datos de las cuales son responsables.

Lo anterior, también incide en la confianza pública en las iniciativas de datos abiertos, al percibirse que los beneficios que se sigue de estos esfuerzos no logran compensar los eventuales peligros a la privacidad y a la protección de los datos personales[53]. Esto podría tener un impacto negativo en la capacidad de las organizaciones de recabar datos de las personas, para el cumplimiento de sus funciones. Además de los daños que pueden surgir si los conjuntos de datos abiertos contienen información personal de manera inapropiada, la preocupación por la privacidad y la protección de datos personales podría llevar a los ciudadanos a tratar de compartir menos datos con los gobiernos[54].

[51] GT29, 2013: 13.

[52] Scassa, 2018: 3.

[53] Cuando los datos se utilizan con fines que van más allá de la razón por la que se recopilaron originalmente, las personas se preocupan porque el contexto social de esos datos ha cambiado, afectando las expectativas de control sobre los mismos. Así, las entidades responsables del tratamiento de datos pueden ganarse la confianza del público mediante una comunicación eficaz y prácticas responsables (Green et al., 2017: 74).

[54] Scassa, 2018: 4. A este respecto, se puede mencionar una amenaza denominada como "descubrimiento en línea", esto es, cuando datos personales aparecen en los resultados de búsqueda en sitios web, a través de los motores de búsqueda de Internet. En el pasado, los registros de información pública estaban típicamente disponibles sólo para aquellos sujetos que visitaban presencialmente los lugares de archivo (Green et al., 2017: 19).

V. ELEMENTOS QUE PERMITEN FACILITAR LA TAREA DE DISPONIBILIZAR LA INFORMACIÓN DEL SECTOR PÚBLICO SIN AFECTAR EL DEBER DE RESGUARDAR LOS DATOS PERSONALES.

1. Principios generales para equilibrar las políticas de datos abiertos con la protección de los datos personales.

Los Principios de Información Justa o "Fair Information Principles" (FIPs, por sus siglas en inglés) proporcionan un marco para equilibrar la protección de datos personales y otros intereses[55]. Una de las versiones más influyentes de los FIPs puede encontrarse en las Directrices sobre la protección de la privacidad y los flujos transfronterizos de datos personales de la OCDE, cuyo alcance se extiende a "datos personales del sector público o privado que, debido a la forma en que se procesan, a su naturaleza o al contexto en que se usan, suponen un peligro para la privacidad y las libertades individuales". Estas directrices contienen los siguientes principios:

a) **Principio de limitación de recogida.** Deben existir límites para la recopilación de datos personales y cualquiera de estos datos deberán obtenerse con medios legales y justos y, siempre que sea apropiado, con el conocimiento o consentimiento del sujeto implicado.

b) **Principio de calidad de los datos.** Los datos personales deben ser relevantes para el propósito de su uso y, para dicho propósito, exactos, completos y actuales.

c) **Principio de especificación del propósito.** La finalidad de la recopilación de datos se debe especificar a más tardar en el momento en que se produce dicha recogida, y su uso se verá limitado al cumplimiento de los mismos.

d) **Principio de limitación de uso.** No se debe divulgar, poner a disposición o usar los datos personales para propósitos que no cumplan con el principio de especificación del propósito, excepto si se tiene el consentimiento del sujeto implicado o por imposición legal o de las autoridades.

e) **Principio de salvaguardia de la seguridad.** Deben emplearse salvaguardias razonables de seguridad para proteger los datos personales contra riesgos, tales como pérdida, acceso no autorizado, destrucción, uso, modificación o divulgación de los mismos.

f) **Principio de transparencia.** Debe existir una política general sobre transparencia en cuanto a evolución, prácticas y políticas relativas a datos personales. Se debe contar con medios ágiles para determinar la existencia y la naturaleza de datos personales, el propósito principal para su uso y la identidad y lugar de residencia habitual de quien controla esos datos.

[55] Zuiderveen Borgesius et al. 2015: 2101.

g) **Principio de participación individual.** Todo individuo tiene derecho a: (i) que el responsable de datos u otra fuente le confirme que tiene datos sobre su persona; (ii) que se le comuniquen los datos relativos a su persona; (iii) que se le expliquen las razones por las que una petición suya haya sido denegada, así como poder cuestionar tal denegación; y expresar dudas sobre los datos relativos a su persona y, si su reclamación tiene éxito, conseguir que sus datos se eliminen, rectifiquen, completen o corrijan.

h) **Principio de *accountability*.** Sobre el responsable del tratamiento de los datos debe recaer la responsabilidad del cumplimiento de las medidas que hagan efectivos los principios señalados anteriormente.

Los datos personales debiesen siempre ser manejados en concordancia con los FIPs, al expresar un consenso global sobre estándares mínimos para el uso justo de esta clase de datos[56].

2. <u>Herramientas técnicas</u>.

Adicionalmente, resulta necesaria la aplicación de soluciones técnicas que permitan, por una parte, identificar previamente los riesgos a la privacidad y a la protección de los datos personales de un proyecto en particular y, a la vez, ofrecer un tratamiento adecuado y con los estándares adecuados de seguridad de los datos e información tratada.

a) **Privacidad por diseño y por defecto.**

Se trata de un principio del derecho a la protección de datos personales, desarrollado para abordar los efectos sistémicos y en constante crecimiento de las tecnologías de la información y la comunicación, y de los sistemas de datos en red a gran escala[57]. Dice relación con la adopción proactiva por parte del responsable del tratamiento de datos personales de medidas necesarias y en forma preventiva, frente a la eventualidad de usos indebidos que pueden afectar a dichos datos. Así, al diseñar una nueva plataforma o modelo de datos abiertos, se debe asegurar, desde el inicio, que los datos personales serán tratados garantizando los derechos que sus titulares tienen sobre éstos.

Este principio entiende que la protección de los datos constituye una parte esencial e integral de la plataforma o herramienta de procesamiento de información que se está desarrollando, resultando esencial para la minimización de posibles riesgos y para fomentar la confianza de los titulares de los datos personales.

En primer lugar, la **privacidad por diseño** se estructura sobre la base de siete principios fundamentales[58]:

[56] Zuiderveen Borgesius et al. 2015: 2107.
[57] Cavoukian, 2011: 1.
[58] *Ibid.* 2.

(i) Proactivo, no reactivo; preventivo no correctivo.
(ii) Privacidad como la configuración predeterminada.
(iii) Privacidad incrustada en el diseño.
(iv) Funcionalidad total – "todos ganan", no "si alguien gana, otro pierde".
(v) Seguridad extremo-a-extremo – protección de ciclo de vida completo.
(vi) Visibilidad y transparencia – mantenerlo abierto.
(vii) Respeto por la privacidad de los usuarios – mantener un enfoque centrado en el usuario.

La privacidad desde el diseño implica, en definitiva, "utilizar un enfoque orientado a la gestión del riesgo y de responsabilidad proactiva para establecer estrategias que incorporen la protección de la privacidad a lo largo de todo el ciclo de vida del objeto (ya sea este un sistema, un producto hardware o software, un servicio o un proceso). Por ciclo de vida del objeto se entiende todas las etapas por las que atraviesa este, desde su concepción hasta su retirada, pasando por las fases de desarrollo, puesta en producción, operación, mantenimiento y retirada."[59].

La aplicación de un enfoque de privacidad por diseño obliga al responsable del tratamiento de datos, de forma previa a que se realice este tratamiento, a determinar las medidas técnicas y de organización que sean necesarias, a efectos de otorgar las garantías de que el tratamiento se realizará en conformidad con las disposiciones legales vigentes. Entre otras medidas, deberá asegurar la confidencialidad de los datos personales, la minimización de los datos, e implementar los mecanismos que sean necesarios para asegurar el pleno ejercicio de los derechos que la ley consagra.

Aquí, cobra relevancia también el empleo de taxonomías depuradas para la categorización y la organización de los datos que son gestionados por el responsable del tratamiento. La taxonomía de datos consiste en la clasificación de diversos tipos de información en diversas categorías y subcategorías, proporcionando una vista unificada de los datos en una organización e introduciendo terminologías y semánticas comunes, aplicables a distintos sistemas de procesamiento de datos. Así, por ejemplo, la taxonomía permite establecer jerarquías dentro de un conjunto de datos, los que son segregados en múltiples grupos, dependiendo de los niveles de cuidado o seguridad que requieren y los riesgos asociados a su tratamiento.

Para traducir en términos prácticos y operativos el principio de privacidad por diseño, resultan de utilidad los procesos sistémicos de ingeniería de la privacidad, el trabajo en

[59] AEPD, 2019a: 6.

estrategias[60] y patrones de diseño de la privacidad[61], junto con la incorporación de tecnologías de privacidad mejorada, que se utilizan para implementar los patrones de diseño de la privacidad con una tecnología concreta[62].

La eficacia está en el centro del concepto de protección de datos por diseño, lo que significa que los responsables del tratamiento deben ser capaces de demostrar que han aplicado medidas específicas para proteger este principio, y que han integrado salvaguardias específicas que son necesarias para asegurar los derechos y libertades de los titulares de datos. Por lo tanto, no basta con aplicar medidas genéricas únicamente para documentar el cumplimiento de este principio, sino que, además, cada medida implementada debe tener un efecto real. Para demostrar el cumplimiento de lo anterior, el responsable del tratamiento puede establecer una serie de indicadores clave de desempeño, incluyendo tanto métricas cuantitativas (niveles de riesgo, reducción de reclamos, reducción de tiempos de respuesta frente al ejercicio de derechos ARCO) como cualitativas (evaluaciones de rendimiento o evaluaciones de expertos[63]).

En cuanto a la **privacidad por defecto**, se refiere a las elecciones realizadas por el responsable del tratamiento en relación con cualquier valor de configuración u opción de procesamiento preexistente, que tenga por efecto ajustar, en particular, pero no exclusivamente, la cantidad de datos personales reunidos (tanto desde un punto de vista cuantitativo como cualitativo), el alcance de su procesamiento, el plazo de almacenamiento (limitando el período de retención a lo estrictamente necesario) y su accesibilidad (limitando quién puede tener acceso a los datos personales)[64]. Este principio requiere que los responsables de las bases de datos sólo procesen los datos

[60] Recoger y tratar la mínima cantidad de datos posible; limitar la exposición de los datos; mantener contextos de tratamiento independientes que dificulten la correlación de grupos de datos que deberían estar desligados; limitar al máximo el detalle de los datos personales que son tratados; informar a los titulares del tratamiento de sus datos en tiempo y forma; proporcionar a los titulares control en relación a la recogida, tratamiento, usos y comunicaciones realizadas sobre sus datos personales; asegurar que los tratamientos de datos personales son compatibles, y respetan los requisitos y obligaciones legales impuestos por la normativa; y, demostrar el cumplimiento de la política de protección de datos que el responsable del tratamiento esté aplicando (*Ibid.* 18 ff.).

[61] Soluciones reutilizables que se emplean para resolver problemas comunes y repetibles de privacidad que se presentan de forma reiterada en un contexto concreto durante el desarrollo de productos y sistemas, v.g. ofuscación de medidas mediante agregación de ruido, agregación en el tiempo, privacidad diferencial, granularidad de ubicación dinámica y controles de acceso selectivo (*Ibid.* 34 ff.).

[62] *Ibid.* 16. Las tecnologías de privacidad mejorada se clasifican, según la finalidad que persiguen, en dos categorías: protección de privacidad, los que incluye productos y servicios para anonimizar, herramientas de cifrado, filtros y bloqueadores (que evitan emails y contenido web no deseado), y supresores de seguimiento (que eliminan las trazas electrónicas de la actividad digital del usuario); y, gestión de privacidad, lo que incluye herramientas de información (que crean y verifican las políticas de privacidad) y herramientas administrativas de gestión de identidad y permisos de usuario (*Ibid.* 28).

[63] CEPD, 2019: 7.

[64] *Ibid.* 10.

que son necesarios para lograr un propósito específico y predeterminado. Para garantizar la protección de datos por defecto es necesario especificar estos propósitos antes de que comience el procesamiento, informándolos adecuadamente a los titulares de datos, y tratar únicamente aquellos datos necesarios para lograr esta finalidad.

Estas medidas se encuentran fuertemente enraizadas en el principio de minimización de datos, tendientes a garantizar que sólo sean objeto de tratamiento los datos personales estrictamente necesarios para cumplir finalidades prestablecidas, con independencia del conjunto de datos recogidos por el responsable. Así, se debiese compartimentar el uso del conjunto de datos entre los distintos tratamientos, de tal forma que no todos los tratamientos accedan a todos los datos. Para dicho fin, al momento de recolectar los datos, estos deben ser clasificados, verificando las características de estos, generando catálogos de datos.

Adicionalmente, se debe considerar elementos como: adoptar un enfoque de "privacidad primero" con cualquier configuración predeterminada de sistemas y aplicaciones; asegurarse de que no se entrega a los individuos una elección ilusoria en relación con los datos que se van a procesar; no procesar datos adicionales, a menos que el individuo consienta en ello; garantizar que los datos personales no se hagan, de forma automática, disponibles públicamente para terceros, a menos que los titulares de datos decidan hacerlo así; y, proporcionar a las personas suficientes controles y opciones para ejercer sus derechos[65]. Así, por ejemplo, la seguridad de la información debe ser considerada como una característica por defecto para todos los sistemas, transferencias, soluciones y opciones en el tratamiento de los datos personales[66].

Tal como se establece en el Reglamento General de Protección de Datos Personales de la Unión Europea (RGPD), la privacidad por diseño y por defecto permite reducir al máximo el tratamiento de datos personales, dar transparencia a las funciones y el tratamiento de datos personales, permitiendo a los interesados supervisar el tratamiento de datos y al responsable del tratamiento crear y mejorar elementos de seguridad. "Al desarrollar, diseñar, seleccionar y usar aplicaciones, servicios y productos que están basados en el tratamiento de datos personales o que tratan datos personales para cumplir su función, ha de alentarse a los productores de los productos, servicios y aplicaciones a que tengan en cuenta el derecho a la protección de datos cuando desarrollan y diseñen estos productos, servicios y aplicaciones, y que se aseguren, con la debida atención al estado de la técnica, de que los responsables y los encargados del tratamiento están en condiciones de cumplir sus obligaciones en materia de protección de datos. Los principios de la protección de datos desde el diseño y por defecto también deben tenerse en cuenta en el contexto de los contratos públicos" (RGPD, considerando 78)[67].

[65] ICO, 2019: 185.
[66] CEPD, 2019: 11.
[67] El artículo 25 del RGPD establece la obligación de protección de datos desde el diseño y por defecto, teniendo especialmente presente el estado de la técnica; los costos de implementación; la

En este sentido, resulta importante llevar a cabo evaluaciones de impacto en la privacidad y en la protección de datos personales, con anterioridad a la disponibilización de información del sector público, en orden a precaver la eventual divulgación no autorizada de datos de carácter personal. La evaluación de riesgos gira en torno a la proporcionalidad: qué daños pueden seguirse a partir de la divulgación de los datos, cuán sensibles son, si están disponibles en otros lugares y qué tan costoso y arduo sería volver a identificar a algún individuo[68]. Con dicha evaluación, se podrán determinar de mejor manera las medidas necesarias para resguardar la privacidad de los datos personales contenidos en las bases de datos objeto de tratamiento.

A este respecto, surge el concepto de "control de calidad de datos abiertos", como una etapa en el proceso de disponibilización de esta información, tendiente a evitar, desde el inicio de las operaciones de tratamiento, la eventual vulneración de la normativa sobre protección de datos personales. En este sentido, un primer paso sería determinar la probabilidad que las bases de datos que se publican contengan datos personales. En segundo lugar, correspondería cotejar el régimen jurídico aplicable a los datos personales en cuestión, luego de lo cual podrá determinarse si esta información puede ser publicada o puesta a disposición del público general en un determinado portal, a partir de la determinación de las reglas jurídicas aplicables (i.e. existencia de una habilitación legal para su tratamiento). Sobre esa base, debiese identificarse aquellos datos que revisten la calidad de "datos públicos".

La implementación de la privacidad por diseño y por defecto resulta altamente beneficiosa para las personas, ya que se garantiza que el tratamiento de sus datos, de manera masiva, se encontrará resguardado por soluciones técnicas que les asegurarán la confidencialidad de su información y, a la vez, beneficia a las organizaciones que realizan grandes tratamientos de datos, ya que *a priori* adoptan las medidas de seguridad respecto de los datos que van a tratar, desde el diseño del respectivo software, plataforma informática, solución tecnológica o procedimiento.

b) Anonimización.

La anonimización puede definirse como una técnica aplicada a los datos personales con el fin de lograr una desidentificación irreversible[69]. La aplicación de técnicas de anonimización constituye un ejemplo práctico de los principios de privacidad por diseño[70], evitando de manera irreversible la asociación de información de carácter personal con un sujeto ya identificado o susceptible de ser identificado.

naturaleza, ámbito, contexto y fines del tratamiento; y los riesgos de diversa probabilidad y gravedad que entraña el tratamiento para los derechos y libertades de las personas.

[68] Simperl, O'Hara & Gomer, 2016: 18.

[69] GT29, 2014: 7.

[70] ICO, 2012: 7.

Estas técnicas implican la creación de una base de datos, a partir del conjunto de datos personales, en el que en el que la información ya no reviste el carácter de personal. Para ello, es necesario reducir el contenido de información de los datos originales, lo que puede hacerse a través de diversos medios: eliminar los identificadores y cuasi-identificadores como los nombres, las fechas de nacimiento y los códigos postales, o agregando datos[71].

La anonimización se presenta como una solución óptima al dilema de cómo pueden reutilizarse los datos personales dentro de los límites legales[72]. Para ello, se deben tener en cuenta medios y factores objetivos, así como los costes, el tiempo y la tecnología necesarios para materializar una posible identificación[73]. Respecto a la irreversibilidad de la desidentificación, es necesario ponderar la razonabilidad de los medios usados, entendido como criterio para evaluar si el tratamiento de anonimización es suficientemente sólido, es decir, si la reidentificación es "razonablemente" imposible, teniendo presente el contexto y las circunstancias particulares de cada caso[74]

Cabe referirse aquí a los datos seudoanonimizados, definidos como aquella información que, sin incluir los datos denominativos de un sujeto, permite (con una probabilidad razonable) identificarlo, al ser cruzada con datos adicionales que figuran por separado (denominados pseudo-identificadores, cuasi-identificadores o identificadores indirectos). En concreto, la seudonimización simplemente, reduce la vinculabilidad de un conjunto de datos con la identidad original del interesado, siendo, en consecuencia, una medida de seguridad útil[75].

En concreto, seudonimizar consiste en sustituir un atributo por otro en un banco de datos, de forma tal que a pesar de que siga existiendo la posibilidad de vincular indirectamente a una determinada persona con el conjunto de datos origen, dicha acción se dificulta. Así, las herramientas de seudoanonimización no permiten la completa disociación de una determinada información con un sujeto identificable, de manera tal que su aplicación sobre una base de datos no sustrae al responsable del

[71] Mediante la agregación, los datos se muestran como totales, por lo que no se muestra información que permita la identificación de ningún individuo. Los números pequeños en los totales a menudo se suprimen por "borrosidad" u omitiéndose por completo. Sus variantes incluyen la supresión de celdas, control de inferencia, perturbación (por ejemplo, agregar o sustraer cantidades a o de los datos cuantitativos, mantener constantes los medios y las desviaciones al tiempo que se modifican los valores potencialmente identificadores), redondeo, muestreo, elaboración de datos sintéticos e informes tabulares (*Ibid.* 52 ff).

[72] Jaatinen, 2016: 5.

[73] Se debe tener presente las dificultades de lograr en algunos casos la anonimización, debido al avance de la tecnología informática y la disponibilidad por doquier de información. Así, uno de los principales factores de riesgo es la creciente cantidad de datos en línea y fuera de línea (GT29, 2013: 13).

[74] GT29, 2014: 9.

[75] *Ibid.* 3. Entre las técnicas de seudoanonimización más utilizadas, cabe destacar el cifrado con clave secreta, función hash, función con clave almacenada, cifrado determinista o función hash con clave con borrado de clave, y descomposición en tokens (*Ibid.* 22 ff.).

tratamiento del deber de dar estricta observancia a los derechos y obligaciones contenidos en la normativa sobre protección de datos personales. En definitiva, si bien la utilización de mecanismos de seudoanonimización sobre los datos personales rara vez, o nunca, hace que las personas sean inidentificables, su utilización puede ayudar a proteger los intereses de la privacidad, al dificultar el reconocimiento de los titulares de los datos[76].

Los procedimientos de anonimización se inician por lo general con un estudio sobre la probabilidad que los datos puedan relacionarse de nuevo con los datos identificativos de la persona, permitiendo la reidentificación. Así, se debe asegurar que sea prácticamente imposible la asociación de los datos que han sido anonimizados con los datos identificativos de la persona afectada[77]. De esta forma, la anonimización conlleva las siguientes etapas:

- evaluar los posibles riesgos que se pueden derivar de la anonimización;
- determinar qué técnicas de anonimización van a ser las más adecuadas; y,
- velar por las medidas de seguridad necesarias para mantener la anonimización[78].

La identificación de riesgos y vulnerabilidades permite determinar el proceso de anonimización que debe aplicarse, es decir, la secuencia de técnicas que debe aplicarse para obtener datos disociados. Estos procesos pueden agruparse en dos familias: la aleatorización y la generalización[79].

(i) La aleatorización consiste en modificar los atributos de un conjunto de datos para que sean menos precisos, eliminando los vínculos que existen entre los mismos y una persona. Esta técnica protege el conjunto de datos del riesgo de inferencia. Dentro de las técnicas de aleatorización destacan:

- la adición de ruido o perturbación aleatoria: modificar los atributos del conjunto de datos (lo que se denomina como "nivel de ruido", donde los atributos se sustituyen por valores aleatorizados), para que sean menos exactos, conservando no obstante su distribución general.
- la permutación: mezclar los valores de los atributos en una tabla (se intercambian los valores contenidos en el conjunto de datos, trasladándolos de un registro a otro), para que algunos de ellos puedan vincularse artificialmente a distintos interesados.

[76] Zuiderveen Borgesius et al. 2015: 2117.
[77] Con todo, se debe tener presente que las técnicas de anonimización pueden afectar sustancialmente a la idoneidad de los datos para determinados fines (Scassa, 2018: 8), llegando incluso a reducir la utilidad de los datos hasta tal punto que la publicación no está realmente justificada (Simperl, O'Hara & Gomer, 2016: 18).
[78] La identificación de los conjuntos de datos que contienen información personal y su preparación para su divulgación mediante la anonimización puede requerir mucho tiempo y recursos. En algunos casos, los recursos gubernamentales disponibles pueden no ser suficientes para la tarea (Scassa, 2018: 8).
[79] CNIL, 2020.

- la privacidad diferencial: generar vistas anonimizadas de un conjunto de datos, añadiendo ruido aleatorio, al mismo tiempo que el responsable del tratamiento conserva una copia de los datos originales. Los subconjuntos de datos anonimizados se entregan a terceros autorizados como respuesta a una consulta concreta[80].

(ii) La generalización implica cambiar la escala de los atributos de los conjuntos de datos, o su orden de magnitud (generalizándolos o diluyéndolos), para asegurar que sean comunes a un conjunto de individuos. Esta técnica, junto con evitar la individualización de un conjunto de datos, limita las posibles correlaciones del conjunto de datos con otros conjuntos de datos. Dentro de las técnicas de generalización destacan:

- La agregación y la k-anonimización: tienen por objeto impedir que un titular de datos personales sea singularizado cuando se le agrupa junto con, al menos, un número k de personas. Para lograrlo, los valores de los atributos se generalizan hasta el punto de que todas las personas acaban compartiendo el mismo valor. En este sentido, la k-anonimización permite cuantificar hasta qué punto se preserva la anonimidad de los sujetos presentes en un conjunto de datos en el que se han eliminado los identificadores, ya sea mediante generalización de los atributos cuasi-identificadores o la eliminación de ciertos registros fuera de rango[81].
- La diversidad l y la proximidad t: la diversidad l extiende la k-anonimización para garantizar que no puedan realizarse ataques por inferencia deterministas, asegurando de que en cada clase de equivalencia, todos los atributos tienen al menos l valores diferentes (limitando la ocurrencia de clases de equivalencia que tengan una variabilidad de atributos escasa). La proximidad t, por su parte, es un perfeccionamiento de la diversidad l, y consiste en crear clases equivalentes que se parezcan a la distribución inicial de los atributos en la tabla (esta técnica es útil cuando haya que conservar los datos lo más próximo posible a los originales, añadiendo una nueva restricción a la clase de equivalencia)[82].

En la determinación de la técnica de anonimización más adecuada, cobra especial importancia ponderar el riesgo de que, una vez que se ha anonimizado un conjunto de datos, se pueda producir una desanonimización de éstos, estimando objetivamente la probabilidad de reidentificación a partir del conjunto de identificadores indirectos[83]. A

[80] GT29, 2014: 14 ff.
[81] AEPD, 2019b: 9.
[82] GT29, 2014: 18 ff. Si se utilizan técnicas de generalización, es fundamental que el responsable del tratamiento no haga uso de un único criterio de generalización, aunque sea para el mismo atributo. La selección de los criterios aplicables debe realizarse según la distribución de los valores de los atributos en la población dada. Cabe señalar que no todas las distribuciones se prestan a la generalización, al no garantizar la variabilidad en las clases de equivalencia (*ibid.* 27).
[83] El nivel de sensibilidad puede determinar el grado de anonimización necesario para que un conjunto de datos pueda ser publicado como datos abiertos. Asimismo, se plantea que, aunque no

este respecto, pueden identificarse tres riesgos clave de la anonimización: singularización (posibilidad de extraer de un conjunto de datos algunos registros que identifican a una persona específica); vinculabilidad (capacidad de vincular como mínimo dos registros de un único interesado o de un grupo de interesados, ya sea en la misma base de datos o en bases de datos distintas); e, inferencia (posibilidad de deducir con una probabilidad significativa el valor de un atributo a partir de los valores de un conjunto de otros atributos)[84].

Cabe tener presente que las técnicas utilizadas para anonimizar información cuantitativa no son generalmente aplicables cuando se busca anonimizar datos cualitativos, como actas de reuniones, transcripciones de entrevistas o videos. A este respecto, resulta necesario aplicar técnicas diferentes, tales como el tarjado de identificadores directos; difuminar las grabaciones de vídeo para ocultar rostros; disfrazar electrónicamente o volver a grabar material de audio; o, cambiar los detalles en un informe (nombres de lugares, fechas, etc.)[85].

Finalmente, dado los riesgos residuales de reidentificación[86], la anonimización no puede asociarse con el concepto de "liberar y olvidar", esto es, proceder únicamente a la publicación de los datos abiertos dejando de controlarlos o sin hacer ningún seguimiento[87]. Compartir y publicar datos requiere una mentalidad de administración por parte de los responsables de su tratamiento, asegurando los recursos y las estructuras institucionales para cumplir con este deber de cuidado en el futuro[88]. Se

es estrictamente información personal, la "información de identificación demográfica" o la "información de identificación de la comunidad", también podría ser considerada para estos efectos como información que requiere de protección (Scassa, 2018: 6).

[84] GT29, 2014: 12. Un marco para las pruebas de reidentificación es el modelo del "intruso motivado". Este enfoque supone la existencia de un sujeto técnicamente competente que cuenta con ciertos incentivos para tratar de identificar al individuo de cuyos datos personales se han derivado los datos anonimizados, de manera tal que las pruebas tienen como propósito evaluar si este sujeto tendrá éxito (ICO, 2012: 22). La organización que realiza una prueba de intrusión motivada evalúa tanto las posibles motivaciones de este sujeto como los métodos a través de los cuales alguien podría lograr esos objetivos. Al centrarse en la persona que está detrás de un ataque de reidentificación y no sólo en los datos en abstracto, el modelo del intruso motivado proporciona un esquema estructurado para determinar cómo se podría utilizar un conjunto de datos abiertos para vulnerar la privacidad individual (Green et al., 2017: 43).

[85] ICO, 2012: 22. Inevitablemente, la anonimización de datos cualitativos puede llevar mucho tiempo, especialmente cuando se trata de tratamientos a gran escala.

[86] GT29, 2014: 27.

[87] Varios ataques de gran repercusión han demostrado que es posible volver a identificar o conocer detalles sobre las personas descritas en las publicaciones de datos, incluso cuando se aplican las técnicas para des-identificar datos o generar estadísticas agregadas (Wood et al., 2016: 5). En la práctica, existe una zona gris muy significativa, en la que entidad que disponibiliza información del sector público podría creer que un conjunto de datos es anónimo, pero un tercero podría todavía ser capaz de identificar al menos algunas personas utilizando, por ejemplo, otra información disponible públicamente, u otra información de la que disponga (GT29, 2013: 13).

[88] Simperl, O'Hara & Gomer, 2016: 12.

debe tener presente que si la agregación o la anonimización de datos personales no se hacen de manera eficaz, se corre el riesgo de que los individuos puedan ser reidentificados a partir de esos conjuntos de datos[89].

[89] GT29, 2013: 12. Además, la anonimización y la reidentificación son campos de investigación activos en los que se publican con regularidad nuevos descubrimientos, de manera tal que se deben evaluar regularmente los riesgos existentes (GT29, 2014: 4).

DOCUMENTOS CONSULTADOS

Agencia Española de Protección de Datos – AEPD (2019a): "Guía de Privacidad desde el Diseño". Disponible [en línea]: https://www.aepd.es/sites/default/files/2019-11/guia-privacidad-desde-diseno.pdf

Agencia Española de Protección de Datos – AEPD (2019b): "Nota Técnica: La K-Anonimidad como medida de la privacidad". Disponible [en línea]: https://www.aepd.es/sites/default/files/2019-09/nota-tecnica-kanonimidad.pdf

Cavoukian, Ann (2011): "Privacy by Design. The 7 Foundational Principles". Disponible [en línea]: https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Comité Europeo de Protección de Datos – CEPD (2019): "Guidelines 4/2019 on Article 25 Data Protection by Design and by Default".___Disponible [en línea]: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf

Commission Nationale de l'Informatique et des Libertés - CNIL (2020): "L'anonymisation de données personnelles". Disponible [en línea]: https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

Conroy, Amy & Scassa, Teresa (2015): "Promoting transparency while protecting privacy in open government in Canada", en Alberta Law Review, Vol 53, N°1, pp. 175-206.

Consejería de Fomento y Medio Ambiente de la Junta de Castilla y León (2019). "Open Data. Publicación y reutilización de Datos Abiertos como iniciativa de Gobierno Abierto en la Administración". Disponible [en línea]: https://datos.gob.es/es/documentacion/open-data-publicacion-y-reutilizacion-de-datos-abiertos-como-iniciativa-de-gobierno

Digital Science (2019): "The State of Open Data 2019. A selection of analyses and articles about open data, curated by Figshare". Disponible [en línea]: https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783#:~:text=A%20formal%20account%20of%20an,any%20other%20type%20of%20information.&text=The%20State%20of%20Open%20Data%202019%20report%20is%20the%20fourth,articles%20from%20global%20industry%20experts.

Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público.

Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público.

Directiva 2019/1024/UE del Parlamento Europeo y del Consejo, de 20 de junio de 2019, relativa a los datos abiertos y la reutilización de la información del sector público.

European Data Portal (2018): "Open Data Goldbook for Data Managers and Data Holders. Practical guidebook for organisations wanting to publish Open Data". Disponible [en línea]: https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

G8 Open Data Charter. Disponible [en línea]: https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex

Green, Ben et al. (2017): "Open Data Privacy", Research Publication N° 2017-1, The Berkman Klein Center for Internet & Society Research Publication Series.

Grupo de Trabajo del Artículo 29 - GT29 (2013): "Opinion 06/2013 on Open Data and PSI Re-use".

Grupo de Trabajo del Artículo 29 - GT29 (2014): "Dictamen 05/2014 sobre técnicas de anonimización".

Halonen, Antti (2012): "Being Open about Data: Analysis of the UK Open Data Policies and Applicability of Open Data", The Finnish Institute in London. Disponible [en línea]; http://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf

Information Commissioner's Office - ICO (2012): "Anonymisation: managing data protection risk". Disponible [en línea]: https://ico.org.uk/media/1061/anonymisation-code.pdf

Information Commissioner's Office – ICO (2019): "Guide to the General Data Protection Regulation (GDPR)". Disponible [en línea]: https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-0.pdf

Instructivo Presidencial N° 005 de 2012, que imparte instrucciones sobre Gobierno Abierto.

Janssen, Katleen (2012). "Open Government Data and the Right to Information: Opportunities and Obstacles", en *The Journal of Community Informatics*, *8*(2).

Jaatinen, Tanja (2016): "The relationship between open data initiatives, privacy, and government transparency: a love triangle?", en International Data Privacy Law, Volume 6, Issue 1, February 2016, pp. 28–38.

Open Data Charter (2015). Principios carta internacional de datos abiertos. Disponible [en línea]:

https://opendatacharter.net/principles-es/

Pagallo, Ugo & Eleonora Bassi (2013): "Open Data Protection: Challenges, Perspectives, and Tools for the Reuse of PSI", en Digital Enlightenment Yearbook 2013, editado por Mireille Hildebrand, Kieron O'Hara y Michael Waidner, IOS Press, Amsterdam, pp. 179-189.

Scassa, Teresa (2018): "Public draft: Open Data & Privacy", en The State of Open Data, editado por Tim Davies, Stephen Walker, Mor Rubinstein y Fernando Perini.

Simperl, Elena, O'Hara, Kieron & Gomer, Richard (2016): "Analytical Report 3: Open Data and Privacy", European Data Portal.

Unidad de Modernización y Gobierno Digital, Ministerio Secretaría General de la Presidencia (2013). "Norma Técnica para Publicación de Datos Abiertos en Chile".

Viollier, Pablo (2017): "El Estado de la Protección de Datos Personales en Chile", Derechos Digitales. Disponible [en línea]: https://www.derechosdigitales.org/wp-content/uploads/PVB-datos-int.pdf

Wiebe, Andreas & Dietrich Nils (2017): "Open Data Protection. Study on legal barriers to open data sharing – Data Protection and PSI", Universitätsverlag Göttingen.

Wood, Alexandra, O'Brien, David & Gasser, Urs (2016): "Privacy and Open Data Research Briefing", Networked Policy Series, Berkman Klein Center Research Publication No. 2016-16.