BIBLIOTECA INFOTEC
VISTO BUENO DE TRABAJO TERMINAL

Doctorado en Ciencias en Ciencia de Datos
(DCCD)

Ciudad de México, a 14 de enero de 2025

UNIDAD DE POSGRADOS
PRESENTE

Por medio de la presente se hace constar que el trabajo de titulación:

"Hacia una nueva teoría y clasificación legal de los datos: un enfoque integral desde los aspectos legales hasta las interacciones de los datos"

Desarrollado por el alumno: **Jersain Zadamig Llamas Covarrubias**, bajo la asesoría de la **Dra. Olivia Andrea Mendoza Enríquez** y el **Dr. Mario Graff Guerrero**, cumple con el formato de Biblioteca, así mismo, se ha verificado la correcta citación para la prevención del plagio; por lo cual, se expide la presente autorización para entrega en digital del proyecto terminal al que se ha hecho mención. Se hace constar que el alumno no adeuda materiales de la biblioteca de INFOTEC.

No omito mencionar, que se deberá anexar la presente autorización al inicio de la versión digital del trabajo referido, con el fin de amparar la misma.

Sin más por el momento, aprovecho la ocasión para enviar un cordial saludo.

Dr. Juan Antonio Vega Garfias
Subgerente de Innovación Gubernamental

JAVG/jah

C.c.p. Mtra. Analy Mendoza Rosales. – Encargada de la Gerencia de Capital Humano. - Para su conocimiento.
Jersain Zadamig Llamas Covarrubias. - Alumno del Doctorado en Ciencias en Ciencia de Datos. - Para su conocimiento.

DCCD

DOCTORADO EN CIENCIAS EN CIENCIA DE DATOS

INFOTEC CENTRO DE INVESTIGACIÓN E INNOVACIÓN EN TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIÓN

DIRECCIÓN ADJUNTA DE INNOVACIÓN Y CONOCIMIENTO
GERENCIA DE CAPITAL HUMANO
POSGRADOS

“Hacia una nueva teoría y clasificación legal de los datos: un enfoque integral desde los aspectos legales hasta las interacciones de los datos”

TESIS
Que para obtener el grado de DOCTOR EN CIENCIAS EN CIENCIA DE DATOS

Presenta:

Mtro. Jersain Zadamig Llamas Covarrubias

Asesores:

Dra. Olivia Andrea Mendoza Enriquez
Dr. Mario Graff Guerrero

Ciudad de México, Enero, 2025.

INFOTEC POSGRADOS

Agradecimientos

Gratias aeternas ago iis qui fuerunt, qui sunt, et qui erunt animae fortibus et mentibus ardentibus, qui in umbris laborantes lucem scientiae ex tenebris ignorantiae extollunt. Vestra devotione et sacrificiis, ignis perpetuus cognitionis alitur, et ego humilem operam meam offero ut haec flamma numquam exstinguatur.

Doy eternas gracias a quienes fueron, quienes son y quienes serán: a las almas valientes y las mentes ardientes, que trabajando en las sombras elevan la luz de la ciencia desde las tinieblas de la ignorancia. Por vuestra devoción y sacrificios, se alimenta el fuego perpetuo del conocimiento, y yo ofrezco mi humilde esfuerzo para que esta llama nunca se extinga.

Tabla de contenido

Índice de figuras

Índice de cuadros

Siglas y abreviaturas

ACLU	American Civil Liberties Union
AEPD	Agencia Española de Protección de Datos
API	Application programming interface
CCPA	California Consumer Privacy Act
DOF	Diario Oficial de la Federación
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EE. UU.	Estados Unidos de América
ENISA	European Union Agency for Cybersecurity
Europol	Oficina Europea de Policía
FRA	Fundamental Rights Agency
GDC NCI	Genomic Data Commons Natuional Cancer Institute
IA	Inteligencia Artificial
IAPP	International Association of Privacy Professionals
ICO	Information Commissioner's Office
INAI	Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales
NIST	National Institute of Standards and Technology
PDPC	Personal Data Protection Commission Singapore
PETs	Privacy-enhancing technologies
PI	Propiedad Intelectual
REDIPD	Red Iberoamericana de Protección de Datos
RGPD	Reglamento General de Protección de Datos Personales
TOGAF	The Open Group Architecture Framework
UE	Unión Europea
UNCTAD	United Nations Conference on Trade and Development
UNECE	United Nations Economic Commission for Europe
WEF	World Economic Forum

Introducción

Los datos son un activo crítico, su valor reside en su obtención, uso, procesamiento y reutilización. Sin embargo, los datos se tratan como una entidad monolítica y no de manera heterogénea, cuya naturaleza depende del contexto de uso. Además, a pesar de que cada día se producen más datos, no existen suficientes datos disponibles para una reutilización innovadora, ni reglas claras, homogéneas y enfoques integrales para su tratamiento. Por lo anterior, el presente proyecto de investigación pretende desarrollar un entendimiento común sobre los conceptos y teorías relacionados con los datos; definir una clasificación de datos según diversos criterios, así como sus respectivos mecanismos de control de acceso; categorizar a los actores y sus roles en el ciclo de vida de los datos; proponer enfoques de mejora de acceso en el intercambio de datos; y demostrar los tipos de intersecciones entre los datos.

La mayor aportación de esta tesis reside en proponer un modelo innovador y flexible que desbloquea el valor potencial de los datos, superando las restricciones y discrepancias de las teorías jurídicas tradicionales. Este modelo establece un marco integral con reglas precisas para clasificar, gestionar y compartir datos, abordando tanto sus dimensiones legales como sus implicaciones prácticas. En un contexto donde los datos son considerados el pilar de la economía digital, la tesis plantea que su tratamiento no debe ser monolítico, sino adaptado a la heterogeneidad y al contexto de uso.

El modelo propuesto enfatiza la importancia de un flujo responsable de datos, asegurando su disponibilidad y protección a través de enfoques reguladores modernos que permitan equilibrar la privacidad, la seguridad y la accesibilidad. Asimismo, busca generar un ecosistema dinámico donde los datos puedan ser reutilizados de manera innovadora, fomentando el desarrollo tecnológico, la competitividad y la equidad en su uso. Este enfoque no solo redefine el tratamiento de los datos en términos normativos, sino que también propone nuevas estrategias para armonizar los intereses de actores públicos, privados y

sociedad en general, garantizando que los datos fluyan de manera confiable y segura sin comprometer los derechos de las partes interesadas.

En el mismo orden de ideas, la presente tesis trasciende las limitaciones actuales al plantear un modelo que, al desbloquear el valor de los datos, los posiciona como un recurso estratégico clave para el avance de la IA. Este marco aspira a consolidarse como una herramienta esencial para una gobernanza ágil y efectiva en un entorno cada vez más interconectado y dependiente de los datos.

Dicho lo anterior, es necesario emplear diferentes enfoques regulatorios y tecnológicos para mejorar la obtención, acceso, intercambio y la reutilización de datos, abordando las implicaciones de sus fundamentos legales y teóricos. Analizando primero, los diferentes tipos de datos y mecanismos de control de acceso. Segundo, los principales tipos de actores y sus roles. Por último, los diferentes enfoques de mejora del acceso, el intercambio de datos y sus tipos relevantes de interacciones.

El eje central de esta investigación radica en la formulación de un nuevo derecho sobre los datos, diseñado para superar las barreras actuales que limitan su aprovechamiento y valor estratégico. Este estancamiento deriva de la ausencia de un marco jurídico coherente y homogéneo que garantice el flujo responsable e interoperable de los datos en un ecosistema global cada vez más interconectado. El planteamiento subyacente de este derecho reconoce la naturaleza multifacética de los datos, abarcando aspectos de privacidad, seguridad, propiedad intelectual y equidad, lo que exige un enfoque integral y adaptativo que se ajuste a los desafíos tecnológicos y regulatorios contemporáneos.

Dirigido a un público diverso y especializado, este trabajo se presenta como una herramienta fundamental para profesionales que interactúan con datos en distintos niveles, incluyendo responsables del tratamiento, científicos de datos, legisladores y desarrolladores tecnológicos. Cada uno de estos actores desempeña un papel esencial en la creación de un ecosistema de datos eficiente y ético, desde la implementación de prácticas de manejo responsable hasta la formulación de políticas y estándares que impulsen la innovación y protejan los derechos fundamentales.

En este sentido, la investigación no solo busca llenar un vacío normativo, sino también inspirar una nueva cultura de gobernanza de datos, donde las oportunidades tecnológicas y los principios legales converjan para maximizar su impacto positivo en la sociedad. Al proponer soluciones que armonicen los intereses de los distintos sectores, este trabajo establece un precedente para la creación de un estrategías que transformen los datos en un recurso accesible y seguro de transmitir.

Puntualmente, el objetivo de la investigación es crear un nuevo modelo de derechos sobre los datos que desbloquee su valor y garantice los principios legales, en el que los datos personales y no personales, incluidos los datos de las empresas, estén seguros y disponibles.

En síntesis, se pretende construir los fundamentos legales de las regulaciones, creando marcos modelo de contextos que permitan el desarrollo de ecosistemas animados, dinámicos y vigorosos, absteniéndose de crear una reglamentación demasiada detallada y estricta, prefiriendo crear un planteamiento ágil de gobernanza que propicie el desbloqueo de valor de los datos en un espacio único; descifrando la propiedad, uso general compartido y el valor de los datos, para así crear un ecosistema de datos tan abiertos como sea posible y tan cerrados como sea necesario.

A continuación, se presenta una figura que abarca todos los temas tratados en esta tesis, funcionando como un mapa general del presente trabajo de investigación.

Figura 1. Teorías de datos, nuevos derechos de datos, tipos de datos, actores en los datos y formas de intercambio en los datos.

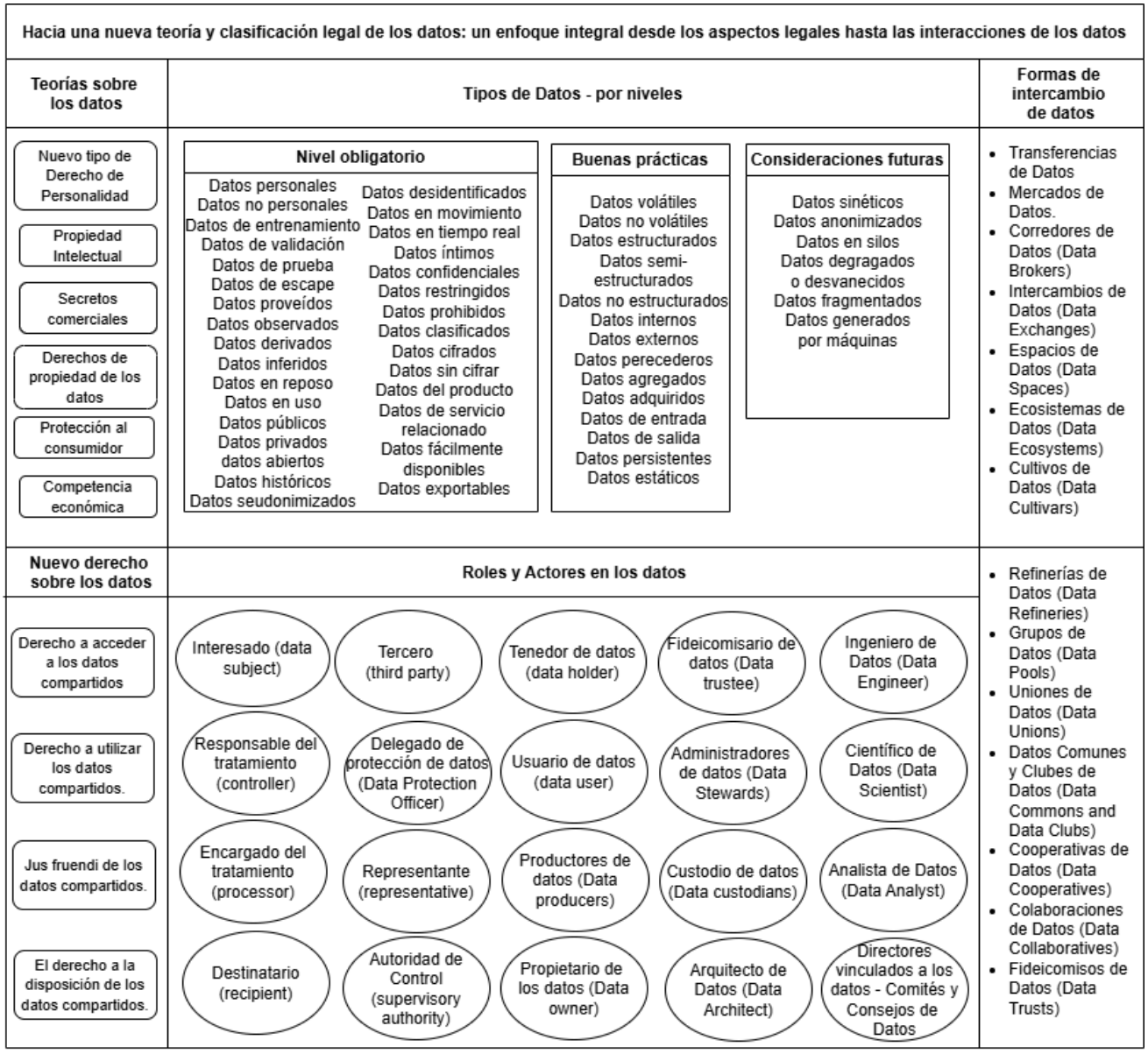

Figura de elaboración propia.

Metodología

El enfoque metodológico adoptado en esta investigación se fundamenta en un análisis comparativo sistemático, estructurado en tres etapas principales: (1)

identificación y selección de normativas internacionales relevantes, con un énfasis particular en las europeas y estándares internacionales; (2) análisis crítico de estas normativas, considerando su aplicabilidad y eficacia en contextos diversos; y (3) evaluación de su posible adaptación a marcos legales, económicos y culturales. Esta metodología permite no solo explorar el potencial de adaptación, sino también identificar vacíos y oportunidades en la legislación mexicana actual.

La elección del análisis comparativo responde a la necesidad de superar la fragmentación normativa en diversos países, especialmente en México, y establecer un marco flexible y adaptativo para la regulación de datos. Este enfoque incluye una revisión exhaustiva de fuentes primarias y secundarias, tales como leyes, estándares, informes regulatorios, y literatura académica. Además, se integraron herramientas cualitativas como el análisis de contenido y el estudio de casos representativos en jurisdicciones que han implementado modelos avanzados de gobernanza de datos.

El modelo propuesto no se limita a una comparación normativa, sino que incluye la construcción de un marco conceptual y metodológico que se ajusta a la diversidad de realidades legales, económicas y culturales. Este marco se desarrolló utilizando una aproximación iterativa, asegurando que la adaptabilidad no comprometa la coherencia normativa ni los estándares esenciales de protección y manejo ético de los datos. Por lo tanto, el modelo combina la uniformidad en principios fundamentales con la flexibilidad requerida para adaptarse a contextos específicos.

Adicionalmente, el marco metodológico integra perspectivas interdisciplinarias para garantizar su relevancia en el contexto global interconectado y sus demandas tan dinámicas. Se consideraron principios de gobernanza de datos, teorías legales contemporáneas y desarrollos tecnológicos emergentes para ofrecer soluciones que promuevan la innovación, la competitividad y la protección de derechos fundamentales.

Para países como México, esta metodología es particularmente pertinente dada la falta de una regulación integral que contemple las demandas del ecosistema digital y las dinámicas del entorno global. Al proporcionar un modelo

práctico y adaptable, esta investigación no solo aborda un vacío normativo, sino que también establece bases para la creación de políticas públicas y estrategias empresariales robustas y confiables.

Capítulo 1

ASPECTOS LEGALES DE LOS DATOS

Capítulo 1. Aspectos Legales de los Datos

Se ha repetido en reiteradas ocasiones que los datos son el petróleo del siglo XXI, pues, así como los combustibles a base de carbono impulsaron la revolución industrial original, los datos están impulsando la cuarta revolución industrial. Sin embargo, independientemente de esta analogía, existen diferencias cruciales entre los datos y el petróleo. Primero, los datos son un producto no exclusivo, es decir, pueden ser utilizados por múltiples actores simultáneamente sin disminuir en volumen o accesibilidad; segundo, cuando los datos incluyen información personal, su recopilación y uso plantean cuestiones de privacidad y otros derechos humanos en formas que no ocurre con el uso del petróleo; (WEF, 2019a, 04) y tercero, los datos no son un recurso finito, ya que los datos engendran más datos, es decir, los conocimientos se pueden derivar continuamente mediante la combinación y el análisis de conjuntos de datos de mayor calidad y cantidad (WEF, 2020a, 13).

Las organizaciones, personas y miles de millones de dispositivos conectados generan, procesan y consumen todo tipo de datos, conllevando a que cada día se produzcan más de 2.5 quintillones de *bytes* de todo tipo de información nueva para ser analizada, procesada y almacenada (OEA, 2019, 05). Esto quiere decir, que el mundo está inundado de datos y crecen exponencialmente, e incluso la mayoría de los datos en el mundo (90%) se han generado sólo en los últimos dos años (WEF, 2021a, 04). Se considera que la esfera global de datos crecerá de 33 *Zettabytes* en 2018 a 175 *Zettabytes* en 2025 (IDC & Seagate, 2018, 03), no obstante, la mayoría de los datos se encuentran en el sector privado, estimando que para el año 2025 el 80% de los datos en todo el mundo residirán en las empresas (WEF, 2020b, 04). Adicionalmente, se estima que el 90% del total de datos no se encuentra en formato estructurado (UNCTAD, 2021, 7).

Aunque el valor de los datos sea intrínseco a su contenido, en una estimación conforme al Cost of Data Breach Report, en 2024, el costo promedio de una violación de datos alcanzó un récord de USD 4.88 millones, un incremento

del 10% respecto al año anterior, impulsado principalmente por la interrupción de negocios y los costos de respuesta post-incidente. La información personal identificable de clientes sigue siendo la más comúnmente comprometida, involucrando el 46% de todas las violaciones, mientras que los registros de propiedad intelectual se encarecieron, alcanzando USD 173 por registro. Los datos almacenados en múltiples entornos fueron los más frecuentemente afectados, representando el 40% de las violaciones, y los datos no gestionados ("shadow data") aumentaron, asociados a un costo adicional promedio del 16.2% (IBM, 2024, 1-18).

Si bien existen brechas de seguridad y costos de las mismas, esto no debe socavar el intercambio de datos, al contrario, se deben tener medidas de seguridad adecuadas a los activos digitales. Ante este advenimiento, sin menoscabo de las medidas de seguridad necesarias, los gobiernos han promovido diversas iniciativas de políticas que mejoran el acceso y el intercambio de los datos, basándose principalmente en cuatro prioridades.

Primero en mejorar el acceso y el intercambio de datos del sector público; segundo, facilitando el intercambio de datos dentro del sector privado; tercero, aumentando las capacidades analíticas de datos en toda la sociedad; y cuarto, desarrollando estrategias nacionales de datos (OCDE, 2019b, 116). Toda vez, que se estima que el acceso e intercambio de datos genera beneficios sociales y económicos por valor de entre el 0,1% y el 1,5% del producto interno bruto (PIB) en el caso de los datos del sector público, y entre el 1% y el 2,5% del PIB (en unos pocos estudios hasta el 4% del PIB) cuando también se incluyen datos del sector privado. La magnitud estimada de los efectos depende del alcance de los datos y del grado de apertura de los mismos (OCDE, 2019a, 11). Estimaciones de la UE muestran que limitar el libre flujo de datos es pérdida de competitividad, conduciendo a una contracción del PIB de la UE de entre el 0,8 y el 3,9 por ciento. El impacto negativo en el ingreso *per cápita* podría ascender a entre $340 y $1,140 (UNCTAD, 2021, 53).

En el mismo sentido, algunas organizaciones como la Unión Europea (UE) y la Organización de las Naciones Unidas (ONU) han expresado su interés en los datos desde una óptica colaborativa por medio de estrategias.

Específicamente la ONU por medio del Secretario General, emitió una Estrategia de Datos para la acción de todos, en todas partes, con conocimiento, impacto e integridad 2020-22, la cual pretende construir un ecosistema integral que desbloquee todo el potencial de los datos, para tomar mejores decisiones en un mayor apoyo a las personas y el planeta, avanzando en la transformación de las personas, cultura, asociaciones, gobernanza de los datos y la tecnología, para que los datos prosperen como un activo estratégico. El resultado de esta estrategia es crear un mejor enfoque de datos para ofrecer resultados superiores en la toma de decisiones más sólidas, asesoramiento sobre políticas, mayor acceso e intercambio de datos, mejor gobernanza y colaboración, protección y privacidad, mayor eficiencia en operaciones, transparencia y responsabilidad en los servicios. La estrategia puntualiza que para desbloquear el potencial crítico se debe centrar en la acción de datos con conocimiento, impacto e integridad en agendas prioritarias clave, cumpliendo los objetivos de desarrollo sostenible o los objetivos globales para el año 2030; la acción climática; igualdad de género; derechos humanos y estado de derecho; paz y seguridad; gobernanza y ética para el futuro y una reforma de la ONU. Por último, determina que se deberá fomentar a los habilitadores que son las personas y cultura; gobernanza y estrategia de datos; alianzas; y entornos tecnológicos (ONU, 2020).

A su vez, la UE por medio de la Comisión Europea, emitió su propia Estrategia Europea de Datos, pretendiendo crear un mercado único de datos que permita a la UE ser más competitiva a nivel mundial, propiciando el desarrollo de procesos, productos y servicios innovadores. Con la creación de dicho mercado único de datos, los datos podrán circular dentro de la UE y entre todos los sectores; se respetarán las normas europeas, específicamente las de privacidad, protección de datos y competencia; así como establecer normas justas, prácticas y claras para el acceso a los datos. Por ello, la UE establecerá normas sobre el acceso a los datos y su reutilización; invertirá en las normas, herramientas e

infraestructura de próxima generación para almacenar y tratar datos; reunirá fuerzas en la capacidad de computación en la nube; pondrá en común los datos de diversos sectores clave; y dotará a los usuarios de derechos, herramientas y capacidades para mantener el pleno control de sus datos. Por último, desde el ámbito regulatorio plantea algunas medidas claves, como un marco legislativo para gobernanza de los espacios de datos europeos comunes; adoptar una norma de ejecución sobre conjuntos de datos de gran valor; y una norma relativa a los datos (Comisión Europea, 2020a).

Dicho lo anterior, el valor de los datos ha cambiado con el tiempo, transmutando de datos hacia conocimiento que ayuda a la sociedad en general y pilar de las economías del futuro. De manera pedagógica se podría clasificar que los datos han transitado por tres etapas diferentes; la primera es la etapa de recursos de datos, donde son un tipo de registro y refleja el mundo real; la segunda es la tapa de activos de datos, donde se consideran más que un recurso, es decir, como un activo que constituye una parte importante de los activos personales o corporativos y la base de la creación de riqueza; y tercera etapa, es la de capital de datos, donde las características de los datos como recurso y activo se desarrollan más y los datos se transforman en capital a través de transacciones y otros métodos de circulación (Yuming, 2021, 15).

Se puede condensar lo dicho hasta aquí, en que el valor de los datos reside en su uso y reutilización. Pero, en la actualidad no hay suficientes datos disponibles para una reutilización innovadora, en particular en el caso del desarrollo de la Inteligencia Artificial (IA). Además, aún hay debates que oscilan en función de quién es el titular de los datos y quién es el usuario de los datos, pero también de la naturaleza de los datos de que se trate (es decir, datos personales, no personales o conjuntos de datos que combinan ambos) (Comisión Europea, 2020a). Estas diferentes implicaciones y condiciones relacionadas con la mayor disponibilidad de datos, desde la producción, gestión y análisis de datos, hasta su potencial en los procesos de toma de decisiones tanto para actores públicos como privados, encuentran síntesis en la expresión de agitación de datos (Concilio &

Pucci, 2021, 04), o incluso como una desprivatización de los datos (Sattler, 2018, 46).

Además, gran parte de los datos que entran en los algoritmos son de naturaleza patentada y no se comparten de manera muy amplia con la investigación, lo que limita la innovación. No obstante, la IA requiere grandes conjuntos de datos para probar y mejorar sus capacidades de aprendizaje, y sin acceso a estos será difícil obtener los beneficios de la IA. Es decir, los datos son el combustible que impulsa el motor de la IA y abrir el acceso a los datos nos ayudará a obtener información que transformará todos los sectores (West & Allen, 2020, 214). Es imperante puntualizar que, los principales elementos que integran la IA son los datos y los algoritmos, y, en consecuencia, sin datos no existe la IA.

En resumen, el presente trabajo de investigación pretende explorar los fundamentos legales relacionados con los derechos sobre los datos, con un enfoque particular en desbloquear su valor y utilizarlos en aplicaciones como en la IA. Esto implica analizar normativas y regulaciones de derecho indicativo (*soft law*) y vinculante (*hard law*) que rigen la recopilación, almacenamiento, procesamiento y compartición de datos. El trabajo de investigación también busca comprender cómo estos marcos legales pueden adaptarse para promover un uso ético y responsable de los datos en el contexto de la IA, facilitando su aprovechamiento para impulsar la innovación y el desarrollo tecnológico, y al mismo tiempo proteger los derechos individuales para garantizar la equidad y la transparencia en su aplicación.

1.1 Desafíos legales y tecnológicos en los Datos

Cuando los datos se comparten, vinculan y combinan a través de fronteras sectoriales e institucionales, se produce un efecto multiplicador, es decir, conectar un *bit* con otro desbloquea nuevos conocimientos y comprensiones que a menudo no se anticipaban (WEF, 2019b, 7). Sin embargo, la tecnología por sí sola no es suficiente, ya que están surgiendo nuevos protocolos, estándares técnicos,

acuerdos y contratos de intercambio de datos para obtener más valor de los mismos (WEF, 2020b, 14), y en este sentido, la regulación ayuda a definir el conjunto de reglas para controlar el acceso y el intercambio de datos, promoviendo la apertura y garantizando el cumplimiento de las normatividades fundamentales, empero, también debe reconocerse, que la regulación puede ser un obstáculo para una buena gobernanza de datos debido a la proliferación de instrumentos fragmentados (OCDE, 2019b, 42).

En consecuencia, las reglas de los datos están incompletas, pues a pesar de que existe un mayor consenso sobre el valor de los datos, a su vez coexisten múltiples teorías jurídicas sobre los fundamentos legales y su concepción, siendo por ejemplo, las teoría de un nuevo tipo de derecho personalidad, teoría de la propiedad intelectual, teoría de los secretos comerciales y teoría de los derechos de propiedad de los datos, y aunque estas teorías e ideas están bien establecidas por las legislación actuales, en la realidad no pueden cubrir todas las características de los derechos de datos (Yuming, 2019, 105-106), principalmente en una sociedad colaborativa con el fin de desbloquear su valor o crear un mercado único.

De seguir así, se tendrá que replantearse los fundamentos de propiedad principalmente sobre el control de los datos, ya sea por los productores, proveedores, consumidores, gobernadores de los datos y la soberanía de los datos. A su vez cuestionarse sobre las bases teóricas del uso general compartido que conlleva abordar la privacidad, seguridad, responsabilidad, confianza, modelado y productos; y por último reconsiderar el valor de los datos, ya sea con beneficios socioeconómicos, utilidad e impacto (Cao, 2018, 121-123). Algunos conceptos erróneos, la falta de claridad sobre conceptos y términos clave continúan saturando el debate de políticas en torno a los datos, su intercambio y la reutilización. Con frecuencia, los datos son tratados como una entidad monolítica o de una sola pieza, pero los avances científicos han evidenciado que son bienes heterogéneos cuyo valor depende del contexto de uso, con diferentes implicaciones para las personas, las empresas y gobierno (OCDE, 2019a, 45).

Actualmente los principios tradicionales de "una propiedad de un objeto" se desafían por "múltiples propiedades para un objeto" y "una propiedad para múltiples objetos", que también se han reconocido indirecta u ocurrentemente reconocido por la ley en cierta medida en prácticas de adjudicación. Las características únicas de los datos, como la reproducibilidad, la inagotabilidad y las publicaciones especiales, hacen posible "múltiples propiedades para uno de los datos". Por lo tanto, permitir que cualquier entidad de datos tenga un control absoluto sobre dichos datos vulnera el principio de compartir (Yuming, 2019, 133).

De la misma manera, ciertamente no se cuenta con soluciones homogéneas de obtención, procesamiento y almacenamiento de datos; no hay suficientes datos disponibles para reutilización, pues a pesar de la disponibilidad de datos abiertos, no son lo suficientemente funcionales, debido a que existen datos que son catalogados como datos personales o intervienen con derechos de propiedad intelectual y/o secretos comerciales. Además, la imprecisión de reglas claras, conlleva a una baja aceptación del intercambio voluntario de datos entre empresas; sumándole que no hay claridad sobre el uso de los datos del sector privado para el bien común; y en consecuencia también una ausencia de enfoques integrales y legales sobre la gobernanza de datos. Por último, a pesar de contar con una regulación en protección de datos, de facto no hay un empoderamiento real del usuario sobre sus datos, principalmente por los mecanismos de portabilidad de datos (Comisión Europea, 2020a).

Por esta razón, un mal manejo de datos o en su defecto, la omisión de una ruta adecuada, pueden exacerbar la desigualdad. A tal efecto, los datos deben fluir de manera fiable y segura con un desbloqueo de valor que respete las leyes de protección de datos, propiedad intelectual, secretos comerciales y propiedad en general. En este sentido, se plantea la necesidad urgente de una clasificación de datos, roles y medios de intercambio que sea armoniosa entre sí, pues sin una comprensión de cómo los datos se posicionan en cada categoría o fluyen entre sí, es imposible tener una discusión sobre cómo gobernar y regular estos flujos.

En el mismo orden de ideas, existe una mayor necesidad de un enfoque más holístico al regular los datos, pero ¿qué debe entenderse como un enfoque

holístico y porque es beneficioso? Un enfoque legal holístico debe alinear diferentes áreas legales a un paradigma de decisión. En la medida que diferentes áreas de la ley sigan sus propios fundamentos y se basen en suposiciones específicas, un enfoque debe acomodar tales subparadigmas o justificar un cambio de paradigma dentro de las áreas particulares. Esto necesariamente requiere un análisis en diferentes niveles de abstracción, mientras que revelan las suposiciones respectivas (a menudo implícitas) y discutiendo de manera sistemática y explícita. Si bien los enfoques holísticos tienen el poder explicativo y analítico a nivel teórico, pueden guiar la regulación para ser tan efectiva como legalmente coherente en un nivel práctico (Richter, 2018, 529).

Pero para llegar a esta visión integrada, se necesitan diversos enfoques para mejorar el acceso, intercambio y reutilización de datos a lo largo de todo el ciclo de vida (OCDE, 2019a, 24), en su obtención, procesamiento, intercambio, análisis y uso (Verhulst et al., 2015), enfocándose primero, en los diferentes tipos de datos y mecanismos de control de acceso, por ejemplo, en diversos datos identificados, seudonimizados, anonimizados, agregados o de dominio personal, privado o público, o control de acceso por las vías de descargas, interfaces de programación de aplicaciones *API* o cajas de arena de datos conocidos como *sandboxes*; segundo, en los principales tipos de actores y sus roles, ya sea por medio de usuarios, intermediarios, repositorios, corredores, mercados, sistemas de gestión, almacenes o terceros de confianza de datos; y tercero, en los diferentes enfoques de mejora del acceso y el intercambio de datos y sus tipos relevantes de interacciones, por medio de acuerdos contractuales y mercados de datos (por publicidad, suscripción, tarifas de uso, venta de bienes o servicios, licenciamiento, tarifas de comisión) o por medio de información abierta por parte de los sectores públicos y privados, y en el caso de particulares en el ejercicio de datos personales, por medio del derecho de autodeterminación informativa y portabilidad de los datos (OCDE, 2019a, 25-44).

Lo dicho hasta aquí justifica la necesidad de mejorar la disponibilidad y gobernanza de un mayor intercambio de datos, no solo publicando más datos abiertos para el acceso abierto, sino también desarrollando licencias y acuerdos

de intercambio de datos (Organización de las Naciones Unidas (ONU), 2020, 30) y en consecuencia, obtener un desbloqueo del valor de los datos en una sociedad colaborativa e implementación de sistemas con IA. Se necesita un marco de gobernanza de datos sólido, compuesto por políticas, leyes, regulaciones e instituciones apropiadas para garantizar que el valor total de los datos se logre y se comparta de manera segura y equitativa (Banco Mundial, 2021, 23).

En síntesis, actualmente los datos se tratan como una entidad monolítica y no de manera heterogénea, cuyo valor depende del contexto de uso. Su concepción legal abarca materias de derecho de personalidad (protección de datos), propiedad intelectual, secretos comerciales, consumidor, competencia económica y propiedad tradicional, que en ocasiones son contradictorias entre sí, y por ende, las reglas sobre los datos están incompletas, ya que en términos de uso, reutilización, procesamiento y soberanía de datos, es necesario considerar un nuevo modelo de derechos sobre los datos que desbloquee su valor y garantice los principios legales.

Una vez que se logre el desbloqueo de valor de los datos, existirán más datos disponibles de origen personal, comercial y gubernamental (no solo datos abiertos), que ayudarán a mejorar múltiples propósitos de la sociedad y los mercados, logrando nuevos hallazgos y soluciones con la combinación de datos con otras tecnologías, verbigracia, con la IA.

1.2 Teorías Legales de los Datos

Los derechos sobre los datos son una preocupación global, con enfoques diversos según el país. En EE. UU., se considera parte de la protección de la privacidad, mientras que la UE lo vincula a los derechos de la personalidad. Japón amalgama el derecho angloamericano con el civil, enfocándose en una comunidad global compartida. China aboga por la cooperación internacional y el manejo adecuado de los desafíos de seguridad en la era del *Big Data* (Yuming, 2020, 228-229).

Con la llegada de la era de los macrodatos, existe un consenso cada vez mayor sobre su valor, pero las reglas de los datos están incompletas (Yuming, 2019, 31). Esto conlleva, a que en la comunidad jurídica se mantengan constantes debates respecto a los derechos sobre los datos y su propiedad, y aunque hasta ahora no se ha llegado a un consenso, las opiniones dominantes generalmente se dividen en diversas teorías. Empero, pese a que estas teorías e ideas están bien establecidas, no pueden cubrir todas las características de los derechos sobre los datos (Yuming, 2019, 105-106).

A continuación, se desglosan las ideas principales de estas teorías desde la concepción de Yuming (2019, 116-119):

1. Teoría de un nuevo tipo de Derecho de Personalidad: se centra en la protección de datos personales generales, privados y sensibles, como nombres o retratos, con un enfoque comparativo que prioriza los intereses individuales de los ciudadanos. Sin embargo, plantea un desafío: los derechos de personalidad son exclusivos e intransferibles, lo que impide considerar cualquier valor económico derivado de ellos como propiedad, ya que esto socavaría el valor de la identidad como persona física.
2. Teoría de los derechos de Propiedad Intelectual (PI): aborda la protección de datos bajo derechos de autor y conexos. Los datos con originalidad pueden ser protegidos como derechos de autor, mientras que aquellos sin originalidad pueden ser protegidos bajo derechos conexos para difundir obras. Sin embargo, identificar qué conjuntos o bases de datos califican bajo la PI es difícil debido a la ambigüedad en cuanto a creatividad y novedad. Además, el valor de la PI se centra en los resultados de la creación intelectual, mientras que los datos sin valor intrínseco radican en su utilización.
3. Teoría de los Secretos Comerciales: sostiene que ciertos datos pueden ser considerados secretos comerciales, ya que su posesión confiere ventajas económicas al titular al evitar su divulgación. Estos datos son exclusivos y su valor radica en su confidencialidad, ya que su revelación a terceros disminuiría su utilidad comercial. El uso del *Big Data* puede permitir la

obtención de secretos comerciales al analizar datos fragmentados, pero esto plantea desafíos, dado que estos datos suelen ser desconocidos para el público y brindan ventajas económicas a sus propietarios. Sin embargo, la inclusión de datos en secretos comerciales puede obstaculizar la libre circulación y aplicación de la información, lo que dificulta la realización de su valor potencial.

4. Teoría de los Derechos de Propiedad de los Datos: postula que cada individuo tiene derecho al valor comercial derivado de sus propios datos personales. En la era digital, estos datos cumplen la función de proteger el interés patrimonial del sujeto. Algunos académicos sostienen que los datos poseen una forma de propiedad, lo que implica el derecho a poseer, utilizar, beneficiarse y disponer de ellos. Sin embargo, surge un inconveniente: si se considera que los datos personales son propiedad, su valor comercial podría exagerarse, lo que podría llevar a descuidar la protección de la privacidad de los ciudadanos. Esto podría afectar la igualdad de personalidad, ya que los datos personales podrían adquirir diferentes valores según las condiciones económicas de las personas. A pesar de esto, es crucial garantizar la protección de la personalidad de cada individuo, independientemente de su situación económica.

Adicionalmente a las teorías anteriores, se demuestra que los datos también encuentran sus límites en los campos de la competencia y el consumidor, volviéndose borroso este lindero conforme a la evolución de la economía digital, particularmente el caso de las normas que rigen las prácticas de los agentes del mercado en relación con la recopilación y el uso de datos personales. Si bien las materias de competencia, consumidores y protección de datos comparten objetivos comunes como proteger al público en general y contribuir al funcionamiento del mercado interior, de facto, los medios por los que persiguen estos objetivos difieren (Graef, 2018, 121-123). Por ejemplo, la ley de competencia tiene como objetivo garantizar la disponibilidad de opciones, mientras que la protección de datos y protección del consumidor deben empoderar individuos para

ejercer eficazmente tal elección, por lo que estas materias siempre deberán ir de la mano para proteger adecuadamente los intereses de las personas y el bienestar del consumidor en general (Graef, 2018, 132).

Dicho lo anterior se agregarían dos tópicos adicionales a las teorías de los datos, siendo:

5. Teoría de los derechos de protección al consumidor: se centra en garantizar que los consumidores tengan control, transparencia, privacidad, seguridad y responsabilidad en el manejo de sus datos personales por parte de las empresas. La diferencia con la protección de datos personales radica en que la teoría de los derechos de protección al consumidor amplía esta perspectiva al considerar los datos como parte de una relación más amplia entre consumidores y empresas, abarcando un espectro más amplio de preocupaciones relacionadas con los derechos y el bienestar de los consumidores en el mercado digital. Un ejemplo entre legislaciones *ad hoc* a este modelo se puede encontrar en EE. UU.
6. Teoría de los derechos de competencia: se centra en promover un entorno competitivo justo y equitativo en el mercado de datos. Esto incluye garantizar un acceso equitativo a los datos, prevenir prácticas anticompetitivas en su recopilación y uso, proteger la privacidad del consumidor y regular fusiones y adquisiciones que involucren grandes volúmenes de datos para evitar distorsiones en la competencia. En resumen, la protección de datos se considera esencial para mantener la competencia justa y proteger los intereses de los consumidores en el mercado digital.

Figura 2. Teorías legales sobre los datos

Figura de elaboración propia.

Un claro ejemplo de las diversas concepciones teóricas y legales sobre los datos se puede esclarecer con juicios importantes en EE. UU. contra *Second Life*; un mundo virtual, desarrollado y creado por la empresa Linden Lab. En el año 2007 se suscitó el litigio *Bragg v Linden Lab*, disputa que radicó en que *Second Life* confiscó ilegalmente la propiedad virtual y negó acceso al mundo virtual al usuario Marc Bragg cuando descubrió cómo adquirir terrenos en el mundo virtual a un precio inferior. Bragg presentó una demanda argumentando que su compra equivalía a la propiedad de las tierras virtuales como si fueran bienes inmuebles reales. Además, otro caso fue *Evans et al. v Linden Lab*, donde los demandantes alegaron que compraron artículos y terrenos virtuales en *Second Life* y unilateralmente se les canceló su cuenta y no fueron compensados por su pérdida.

Sin embargo, los tribunales jamás entraron al fondo del asunto y el debate permanece en los límites de la academia (Llamas Covarrubias, 2020, 276-278).

Considerando los casos mencionados anteriormente, se puede demostrar cómo diversas teorías de los datos pueden ser contradictorias entre sí. Verbigracia, se podría alegar que los términos y condiciones de la plataforma no otorgan propiedad desde la *teoría del consumidor*, o que los datos (propiedades virtuales) eran acorde a la *teoría de propiedad intelectual* por haber delegado licencias a la plataforma, o como en la *teoría de la propiedad tradicional*, extendiendo la propiedad real corpórea a la incorpórea y delegando el título de bienes absolutos a las propiedades virtuales. También podría plantearse hipotéticamente, que Bragg o Evans podrían recurrir a la *teoría de personalidad*, específicamente de protección de datos, si algún delincuente informático o desarrollador sin permisos tomó información o conversaciones importantes de los usuarios, así como la suplantación de un avatar digital. De manera más particular, en la circunstancia de robo de información importante que otorgue ventaja competitiva de corporaciones que venden sus productos desde ese mundo virtual, podría vincularse a la *teoría de secretos comerciales*. Incluso, podría considerarse el caso teórico, donde *Second Life* puede fusionarse con otros mundos virtuales (desde una fusión corporativa), como Meta (metaversos) y tenga que recurrir a la *teoría de los derechos de competencia*.

1.3. La discrepancia entre las teorías de los datos

Una vez abordadas las diferentes teorías de los datos, es necesario desentrañarlas y contraponerlas para consolidar una nueva teoría de datos, o en su defecto, redefinir su visión. Respecto al primer postulado, se tiene la **teoría de un nuevo tipo de derecho de personalidad,** que de facto se materializa en los derechos de protección de datos personales. Esta requiere que las leyes deban adaptarse para proteger la información personal en la era digital. Esto implica reconocer que los individuos tienen derechos sobre sus datos, incluido el control sobre su recopilación, uso e intercambio, así como el acceso, rectificación,

cancelación, oposición y portabilidad de esa información. Esta idea ha llevado a la promulgación de leyes de protección de datos como el RGPD en la UE, que busca equilibrar la protección de la privacidad individual con la circulación de la información para propósitos legítimos.

Cabe destacar que, los derechos de personalidad son un conjunto de derechos fundamentales inherentes a la persona, cuyo objetivo principal es proteger su identidad, dignidad, y esfera privada. Estos derechos incluyen, pero no se limitan, al derecho al nombre, la imagen, la voz, la privacidad, y la protección de datos personales. A diferencia de los derechos patrimoniales, los derechos de personalidad son inalienables, intransferibles e imprescriptibles, lo que significa que no pueden ser cedidos, vendidos o renunciados, ni siquiera por la propia persona (Figueroa, 1998, 22). No obstante, esto no implica que no puedan coexistir con la explotación económica de ciertos aspectos relacionados, como el uso autorizado de la imagen o la voz en contextos comerciales, siempre que no se vulnere la esencia de la dignidad personal ni se desnaturalice el núcleo de estos derechos. Por lo tanto, cualquier referencia a los derechos de personalidad debe enfocarse en su carácter protector, sin asumir que estos puedan transformarse en bienes económicos susceptibles de propiedad o comercio.

No obstante, encasillarse únicamente en la protección de datos personales puede ser limitativo para una teoría de los datos. Es decir, tratar esta teoría como la base, debe ser con cautela, toda vez que se podría objetar la comercialización de los derechos de la personalidad, temiendo que una emancipación de los datos personales pueda tener un impacto negativo en la libertad de expresión y comunicación social. Incluso si se constituye ese derecho a los datos personales, continuará siendo un desafío asignar derechos de acceso, uso e intercambio de datos personales a costos de transacciones razonables (Sattler, 2018, 43).

Por otra parte, desde la concepción de la segunda **teoría de los datos de propiedad intelectual**, se postula que los datos son activos valiosos que pueden ser poseídos, controlados y monetizados como otros activos de propiedad intelectual. Se podría argumentar que quienes generan o recopilan datos deben tener derechos legales para proteger y gestionar esos datos, incluido el derecho a

determinar su uso, compartirlos y obtener beneficios económicos de ellos. Esta perspectiva ha dado lugar al desarrollo de leyes y regulaciones relacionadas con la propiedad de datos, la privacidad y la seguridad de la información, aunque plantea desafíos éticos y de acceso a la información.

Específicamente esta teoría presenta desafíos en su aplicación práctica. Desde una perspectiva maximalista, cualquier obra que requiera cierto esfuerzo o habilidad podría estar sujeta a derechos de autor, incluso hojas de cálculo simples. Sin embargo, una visión más moderada requiere demostrar expresión creativa para obtener protección, lo que excluye elementos funcionales como los sistemas operativos según los tribunales alemanes. Además, la necesidad de demostrar la originalidad de los datos plantea problemas, especialmente en entornos como la computación en la nube, donde muchos datos se generan automáticamente. En lugar de centrarse en la propiedad intelectual de los datos en sí, podría ser más adecuado enfocarse en los derechos para controlar la reproducción y la comunicación pública (Reed, 2021, 179-181).

En el mismo orden de ideas los derechos de propiedad intelectual son extremadamente variados. Las patentes protegen las invenciones, los derechos de diseño protegen los diseños decorativos y funcionales, las marcas protegen los signos que indican el origen de los bienes y servicios, los derechos de ejecución protegen la forma en que se realizan las obras, etc. Cualquiera o todos estos derechos pueden ser aplicables a información particular en la nube, pero su enfoque principal no es la información en sí, sino algún otro aspecto de la cosa protegida. Así, por ejemplo, una patente confiere el derecho a impedir que otros fabriquen o utilicen la invención; aunque la patente se describe utilizando información, esa información no está en sí misma protegida por el derecho (Reed, 2021, 179).

Además, en relación con las bases de datos con datos personales procesadas con fines comerciales, no es una cuestión clara si la inversión radica en la creación o en la recopilación, ya que técnicamente los controladores de datos no crean datos personales, en su lugar los recopilan de las personas. En esta interpretación, la lista de clientes y grupos de correo electrónico (datos

procesados con fines de *marketing* directo) son datos recopilados en lugar de creados por el controlador de datos, pero a su vez se requiere recopilar el consentimiento de los usuarios y proporcionar mecanismos de obtención, cancelación, verificación y actualización (Banterle, 2018, 425).

En tercer lugar, respecto a la **teoría de los datos sobre la protección del secreto comercial o industrial**, se sostiene que la información estratégica de una empresa, como fórmulas, procesos y clientes, debe mantenerse en secreto para preservar su ventaja competitiva. Esto implica establecer medidas legales y prácticas para evitar su divulgación no autorizada, como acuerdos de confidencialidad y políticas internas de seguridad de la información. Proteger estos datos es crucial para la innovación y competitividad de la empresa, evitando que sean aprovechados por competidores o divulgados públicamente sin consentimiento.

No obstante, a pesar de que el secreto industrial pertenece en *stricto sensu* al marco jurídico de propiedad intelectual, para fines materiales, la naturaleza de este activo no necesita registro y por ello se separa de aquellos activos declarativos y/o constitutivos de la propiedad intelectual. Materialmente, el secreto industrial es cualquier tipo de información de la empresa que otorgue ventaja competitiva y/o económica y que se tenga un nivel de secrecía esta, específicamente puede incluir listas de clientes o correos electrónicos, ya sean reales o potenciales, técnicas de *marketing* y conjuntos de datos de clientes perfilados, políticas de precios, descuentos, y datos relacionados con promociones y venta de productos.

Cabe destacar que las entidades también poseerán información que no es un secreto comercial pero que también es clasificada como confidencial (por ejemplo, la razón por la que se despidió a un empleado) y las entidades que no pertenecen a la entidad, como las personas físicas o las organizaciones benéficas, también poseerán información confidencial (Reed, 2021, 183). Esto quiere decir, que el término “confidencial” es polifacético, pues una cosa es el deber de confidencialidad en los datos personales que están vinculado al de seguridad; otra cosa es la confidencialidad de la información en sí misma, aunque no sean datos

personales *per se*, y otra es la confidencialidad vinculada a la secrecía de los secretos industriales y/o comerciales.

En cuarto lugar, desde la interpretación de los datos desde **la teoría de propiedad,** se postula que los datos son activos con un valor intrínseco y que quienes los generan o recopilan deben tener derechos legales sobre ellos, es decir, se refiere a la idea de considerar los datos como un tipo de propiedad que puede ser poseída, controlada y transferida de manera similar a otros activos tangibles o intangibles.

Los objetos corporales e incorporales pueden estar sujetos a derechos de propiedad, pero desde una concepción de teoría sobre los datos existen pequeñas variaciones a esta concepción, respecto a que los datos incorpóreos deben considerarse como la llamada *res digitalis*, porque cumplen los requisitos de controlabilidad y corporeidad impuestos por las leyes de propiedad. Es decir, esta sería una interpretación ampliada de la noción de *res*, incluido los datos digitales, sin embargo, bajo las características de ser bienes no rivales, de naturaleza tangible y por ubicuidad de los datos digitales, este intento de establecer derechos de propiedad con respecto a ellos, causa problemas conforme a los principios fundamentales del derecho de propiedad (Chrobak, 2018, 260).

El principio de especificidad presenta un desafío al determinar el alcance exacto del derecho de propiedad en datos digitales, ya que es difícil distinguir entre el original y las reproducciones. Esto se manifiesta especialmente en la transferencia de propiedad, donde el cesionario puede obtener copias mientras el cedente retiene el control del original, dando lugar a disputas legales complejas sobre la aplicación de estos derechos. Además, la publicidad de estos derechos resulta cuestionable, dado que no pueden ser fácilmente reconocidos por terceros a través de la posesión o el registro. En vista de esto, el derecho de propiedad en su forma actual no es suficiente para garantizar el control de los datos digitales por parte de los individuos, lo que ha generado la necesidad de explorar enfoques alternativos, como los derechos de propiedad virtual en datos digitales (Chrobak, 2018, 261). Blockchain emerge como una solución ante bienes rivales y no rivales, pero eso se abordará en otro apartado.

En quinto lugar, en **la teoría de los derechos de protección al consumidor** en relación con los datos implica garantizar que los consumidores tengan control sobre sus datos personales. Esto incluye derechos como la privacidad de los datos, el conocimiento sobre cómo se usan, la corrección de información incorrecta y la eliminación de datos innecesarios. Pero específicamente, los consumidores deberían tener el derecho a que sus datos no sean vendidos sin su consentimiento y a no ser discriminados por no proporcionar ciertos datos. Estos derechos buscan proteger la privacidad y seguridad de los individuos en un mundo digitalizado y asegurar que las prácticas comerciales sean justas y transparentes.

Si bien a *prima facie* se pudiera encontrar una intersección entre las teorías de los **derechos de personalidad materializada en la protección de datos personales y la teoría de los derechos de protección al consumidor**, estas formalmente cuentan con diferencias esenciales. Al respecto Ratti (2018, 146) cuestiona y expresa ¿El retiro del contrato y del consentimiento tienen el mismo fundamento? Los consumidores participan en contratos vinculantes, mientras que los titulares en la protección de datos personales entran en la definición de derechos fundamentales y estos deben procesarse de manera justa para fines específicos y sobre la base del consentimiento de la persona interesada o de alguna base legítima legal. Dicho esto, el derecho en protección de datos es catalogado como fundamental y como tal, no puede ser objeto de ningún contrato, y por tal motivo, el derecho de desistimiento o retirada del consentimiento por el interesado previstas en la normatividad de protección de datos no tiene límite de tiempo.

La contraposición entre la teoría de protección de datos personales, que podría fundamentarse en normativas como el RGPD (Reglamento General de Protección de Datos), y la teoría de protección al consumidor, que se podría a asociar al marco normativo CCPA (*California Consumer Privacy Act*), plantea desafíos debido a sus enfoques divergentes. Mientras que la RGPD se centra en salvaguardar la privacidad y el control de los individuos sobre sus datos personales, la CCPA está más orientada a garantizar la equidad en las

transacciones comerciales y la transparencia en la recopilación y uso de datos personales por parte de las empresas. Esto puede generar conflictos en términos de requisitos de consentimiento, derechos de acceso y rectificación, así como en la imposición de sanciones por incumplimiento. Además, la aplicación simultánea de estas regulaciones puede generar complejidades y confusión tanto para las empresas como para los individuos, especialmente en un entorno tecnológico en constante evolución.

Finalmente, respecto a la última teoría; la **teoría de los derechos de competencia**, en el contexto de los datos busca garantizar una competencia justa y equitativa en el mercado de datos. Esto implica prevenir prácticas anticompetitivas como la concentración excesiva de datos en manos de unas pocas empresas, el abuso de posición dominante para controlar el acceso a datos importantes, y acuerdos colusivos que limiten la competencia. Al promover una competencia saludable en el mercado de datos, se busca proteger la innovación, fomentar la elección del consumidor y promover un entorno económico en el que todas las empresas puedan competir en igualdad de condiciones para beneficiar a los consumidores y la sociedad en general.

Un ejemplo es la fusión entre *Microsoft* y *LinkedIn* en Europa. En este caso, se evidenció que la elección del consumidor en cuanto a la protección de la privacidad se convirtió en un factor crucial de competencia entre las redes sociales profesionales. Se argumentó que esta fusión podría amenazar la competencia si el mercado favorecía notablemente a *LinkedIn*, lo que demostró que la cuestión en juego trascendía el simple análisis de competencia basado en el precio. La discusión se centró en si la ley de competencia debía emplearse para promover un nivel más alto de protección de datos. Así, la protección de datos emergió como un criterio determinante en la competencia entre empresas en un mercado específico (Graef, 2018, 125).

A este respecto, en 2012, el *Bundeskartellamt* alemán determinó que una empresa dominante podría violar las leyes de privacidad al recopilar datos personales sin consentimiento para obtener ventajas sobre competidores. Otro caso importante fue el de Facebook; en 2016, *Facebook* fue objeto de sospechas

por abuso de su posición dominante en redes sociales, vulnerando la protección de datos y imponiendo términos injustos a usuarios. En 2017, se concluyó preliminarmente que Facebook abusaba de su posición al requerir la recopilación de datos de terceros para usar su plataforma, sin opción de elección para los usuarios (Graef, 2018, 139).

En EE. UU., es importante el caso de la fusión *Google/DoubleClick*, donde Pamela Jones Harbour, ex comisionada de la Comisión Federal de Comercio de los EE.UU., sugirió establecer un cortafuego entre los datos de *Google* y DoubleClick durante un tiempo determinado, que buscaba prevenir posibles efectos anticompetitivos. Cuando la combinación de conjuntos de datos se perciba como una amenaza significativa para la competencia efectiva, podría requerirse que las partes fusionadas mantengan sus bases de datos separadas por razones de eficiencia económica. Es crucial destacar que este enfoque también impediría el uso indebido de datos personales para fines que no cumplan con la legislación de protección de datos (Graef, 2018, 146).

En síntesis, estas teorías sobre los datos abordan aspectos diversos, desde la protección de la privacidad individual hasta la regulación de la competencia en el mercado de datos. La teoría de un nuevo tipo de derecho de personalidad y los derechos de protección al consumidor se enfocan en proteger los datos personales y los derechos de los consumidores, respectivamente. Por otro lado, la teoría de los datos de propiedad intelectual y la teoría de propiedad tratan los datos como activos sujetos a derechos de propiedad. La teoría de los datos sobre la protección del secreto comercial se centra en resguardar datos empresariales estratégicos, mientras que la teoría de los derechos de competencia busca garantizar una competencia justa en el mercado de datos. Estas teorías pueden crear sinergias en aspectos como la protección de la privacidad y la promoción de la competencia, pero enfrentan retos en términos de implementación efectiva y resolución de conflictos entre diferentes intereses y derechos.

En lo general, si bien la información personal opera como una moneda, y en ocasiones la única moneda en el intercambio de servicios en línea. En la práctica, mediante técnicas invasivas de intercambio de datos, las empresas son

responsables de hacer un mal uso de los datos personales para obtener una ganancia económica, es decir, los datos personales (de esfera privada) se convierten en datos públicos propiedad de terceros para que las empresas dominantes puedan consolidar aún más su posición dominante (Chirita, 2018, 157-158).

En particular, si se toma un enfoque patentado de información general, se corre el riesgo de restringir el acceso al conocimiento, sofocando la innovación y el progreso. La creación de un monopolio legal sobre los datos corre el riesgo de generar una protección excesiva, pues en el contexto de los macrodatos, la reutilización, enriquecimiento y acceso a datos son esenciales. Por lo tanto, deberá garantizarse el acceso a los datos y la creación de conocimiento, evaluando cuidadosamente la introducción de fuertes derechos exclusivos sobre los datos (Banterle, 2018, 438). Sin embargo como antítesis, también puede inferirse que los desafíos legales y económicos confunden la definición de los derechos de propiedad sobre los datos, involucrando una ambigüedad entre los derechos de propiedad del interesado y del recolector de datos, teniendo este último conforme a las normatividades en protección de datos; un derecho de propiedad puramente de facto, que si bien beneficia la protección de datos personales, desde otro enfoque advierte el detrimento de una baja asignación eficiente de los derechos de propiedad, imposibilitando compartir datos (Banco Mundial, 2021, 32).

Además, algunos académicos asocian la propiedad de los datos personales con una mayor seguridad jurídica, pero otros han criticado este enfoque como una visión mítica de la propiedad privada, temiendo que un derecho de propiedad pueda sofocar el desarrollo de la infraestructura y los servicios necesarios para acceder a los datos personales, y por lo tanto, bloquear la investigación e innovación. Incluso se ha demostrado que el derecho a los datos personales no puede basarse únicamente en contratos consensuales, sino que requeriría limitaciones legales otorgando acceso no consensuado también (Sattler, 2018, 46-47). Aunado a que con el surgimiento de las licencias cruzadas donde un usuario realiza una creación mixta con información licenciada por un tercero o

proveedor no infringe los derechos de los terceros, pero esto no significa que el usuario creador sea propietario de la creación mixta (Reed, 2021, 189).

1.4. Hacia una nueva teoría de datos

Llegados a este punto, una vez analizadas las teorías mencionadas con anterioridad, se puede inferir que estas son razonables hasta cierto punto, pero no agotan todas las circunstancias que deberían ser cubiertas por un verdadero derecho sobre los datos.

Previo analizar una nueva concepción de derechos sobre los datos, surge la pregunta ¿Hacia dónde se debe de transitar para crear una nueva teoría sobre los datos? Se podría argumentar que el derecho sobre los datos es distinto de los derechos de personalidad, secretos comerciales y propiedad intelectual, y debe considerarse como un nuevo tipo de derecho independiente, posiblemente denominado como el derecho sobre activos de datos.

Este derecho se podría describir como una forma de propiedad incompleta, ya que los titulares o responsables del tratamiento no pueden disponer de los datos que poseen o controlan de manera arbitraria, ni dañar los derechos e intereses legítimos de los titulares. Por tanto, es crucial categorizar y establecer un sistema de derechos basado en las diversas formas de existencia de la información. Los derechos sobre los activos de datos implican el control directo de una propiedad de datos específica y la exclusión de la interferencia de terceros, representando una nueva forma de derecho de propiedad emergente en la era de los macrodatos. Estos derechos otorgan al propietario el poder de poseer, usar, obtener beneficios y disponer de sus datos, lo que implica que no son estáticos ni unidireccionales (Yuming, 2020, 78). Otro derecho propuesto por el Banco Mundial (2021, 313), es el derecho a la producción de datos, advirtiendo y haciendo un llamado a legislaciones y marcos regulatorios que protejan los derechos de las personas a producir, usar y difundir datos.

Otros autores se cuestionan si debe surgir un derecho independiente denominado derecho de producir datos u otro tipo de derecho exclusivo sobre los

datos, logrando un equilibrio entre el derecho exclusivo y otros derechos de legislación en protección de datos, derechos de autor, patentes, bases de datos, protección al consumidor, derecho de competencia, derecho privado e incluso derecho penal, proponiendo la conveniencia de clasificar el derecho del productor de datos como un derecho complementario en relación a la mayoría de los campos legales, es decir, la sociedad se encuentra en un momento donde la debe preguntarse si realmente se necesita un derecho exclusivo sobre los datos o se debe considerar la protección indirecta que ya brindan otras instituciones y figuras jurídicas (Steinrötter, 2020, 285-287).

A *contrario sensu*, Sattler (2018, 39-46), se pregunta si es necesaria la transición del consentimiento a la propiedad, ¿tabú o solución? ¿Hacia un derecho de propiedad sobre los datos personales? ¿Hacia una ley de protección de datos dualista?, ya que aboga por el empoderamiento de los interesados al introducir el derecho a los datos personales. Específicamente que esto daría como resultado un derecho de carácter dual, ya que sería similar a una propiedad, al tiempo que integra aspectos fuertes de un derecho de personalidad, argumentando que su teoría de derecho a los datos personales podría tener ventajas para consumidores digitales, ya que aumentaría la conciencia de ofertas digitales y de la contraprestación contractual de un mercado de plataforma lateral. Además, esto podría facilitar la transparencia de las cláusulas contractuales debido a la necesidad de acordar explícitamente una licencia de datos. Si bien esto no resolvería automáticamente todos los sesgos de comportamiento que impiden el proceso de consentimiento, pero claramente, la introducción de un derecho a los datos propios reforzaría el principio de autonomía y aliviaría en parte la carga regulatoria.

Teniendo en cuenta estas consideraciones y aclaraciones, y al eliminar cualquier incertidumbre para establecer una teoría sólida, se toma como base la visión de Yuming (2019, 119-124), que propone un nuevo derecho sobre los datos que no se identifican simplemente con los derechos de personalidad o de propiedad, sino que se asemejan más a una integración de los dos sin dejar de ser independientes, personificando así, los atributos de personalidad y propiedad. Por

otra parte, existen datos con valor económico que pueden considerarse objetos de derecho de propiedad de datos y sin jugar un papel en la protección de datos personales, y por ende solo estar relacionados con los derechos de propiedad.

Pero en el caso de datos personales, si estos sólo se protegen en la categoría de derechos de personalidad, se tratará únicamente de personalidad y dignidad, inhibiendo el flujo de datos y el desarrollo de la industria de datos hasta cierto punto. Por lo tanto, confirmar la naturaleza del derecho de propiedad de los derechos sobre los datos puede ayudar a regular mejor las transacciones de datos, proporcionando un entorno ordenado con respaldo legal sólido para el mercado de datos (Yuming, 2020, 76-77).

Este nuevo derecho sobre los datos se basa en cuatro novedades; siendo 1. Involucran un nuevo tipo de objeto, 2. Son un nuevo tipo de derechos, 3. Representan una nueva naturaleza de los derechos, y 4. Contienen un nuevo tipo de funciones, las cuales se describen a continuación (Yuming, 2020, 57-59). En seguida, se resume cada novedad de este nuevo tipo de derecho:

1. Objeto de derecho: No son un objeto en sentido de la ley civil (no son un objeto de derecho de propiedad), no son un objeto físico ni intangible como se describe en los derechos de propiedad intelectual. Propone que los datos son un objeto independiente y forman un mundo digital más allá del mundo tangible o intangible. Los sujetos de los derechos de los datos son titulares de derechos específicos, incluidos las personas físicas, jurídicas u organizaciones no jurídicas, variando el alcance de los derechos dependiendo el sujeto. Estos derechos son conjuntos de datos específicos, es decir, son una combinación de una serie de dígitos, códigos, imágenes y textos, donde estos últimos no tienen valor hasta que pueden ser combinados, integrados o intercambiados.
2. Tipo de derecho: Según la interpretación jurídica, los derechos humanos se pueden clasificar en derechos de personalidad y derechos de propiedad. Pero en la era digital, los seres humanos dejan huellas en diversos ecosistemas digitales, es decir, por una parte, estos datos son fragmentos de comportamientos humanos: un registro importante de la participación de

las personas en actividades sociales y formas de extender la personalidad. Por lo tanto, es necesario salvaguardar la dignidad humana, así como los derechos a la libertad, reputación, privacidad e información. Por otro lado, los datos también son recursos sociales importantes; por ende, tienen valor y pueden aportar beneficios económicos a los interesados, de manera que es necesario otorgar derechos sobre los activos de datos, y en consecuencia los derechos sobre los datos contendrán tanto derechos de personalidad como derechos de propiedad.

3. Naturaleza del derecho: Los derechos de datos son una combinación de poder público y derechos privados. Incluyen tanto la soberanía de los datos que refleja la dignidad nacional, como el derecho a los datos que resalta el bienestar personal de los humanos, por lo que los derechos sobre los datos deben ser analizados, no solo como una especie de derechos personales regulados por el derecho privado, sino también desde la perspectiva de seguridad del estado regulada por el derecho público. Es decir, los derechos sobre los datos necesitan tanto la autorregulación del derecho privado como la interferencia del derecho público.
4. Función de derecho: Los derechos de propiedad actuales son exclusivos como un tipo de propiedad y nunca puede haber dos propiedades sobre un objeto y todos tienen la obligación de no interferir con el dominio absoluto de los titulares de derechos sobre los objetos de los derechos. Sin embargo, los derechos sobre los datos, por el contrario, no son exclusivos, es decir, son derechos para ser compartidos. Esto se manifiesta por la propiedad múltiple de los datos (que es el núcleo), ya que los derechos para compartir, en esencia, será un hito que reescribe las reglas de la civilización humana.

Ante esto, Yuming (2020, 60) propone un nuevo derecho sobre los datos, destacando que el derecho sobre los datos tiene características de derechos privados, públicos y soberanía. En el ámbito de los derechos privados, se pueden subdividir a datos personales y derechos de datos corporativos. En el ámbito de

los derechos públicos, estos muestran características públicas y colectivas, con la maximización del interés público como valor, manteniendo vigorosamente la participación en asuntos públicos. En el ámbito de la soberanía, los derechos sobre los datos se manifiestan enriqueciendo la connotación de la soberanía estatal tradicional en armonía con una nueva gobernanza del ciberespacio moderno.

Simultáneamente Yuming (2020, 148-149) divide los derechos sobre los datos en cuatro partes, siendo el derecho a acceder a los datos compartidos, el derecho a utilizarlos, el *Jus fruendi* (expresión latina que significa el derecho del propietario de percibir los frutos de su cosa), y el derecho a la disposición de los datos compartidos. Estas partes son desglosadas a continuación (Yuming, 2020, 153-158):

1. Derecho a acceder a los datos compartidos: Significa que las personas tienen derecho a adquirir datos conforme a las leyes y regulaciones, pero está sujeto a la voluntad de las personas que ofrecen datos para compartir, pues nadie puede acceder a los datos sin este requisito. En términos generales, los objetos cubiertos por este derecho pueden incluir recursos de datos, tecnologías, canales y servicios de datos que se comparten. En sentido estricto, solo se comparten datos que son públicos, renovables, urgentes, valiosos y escasos por naturaleza. Asimismo, el acceso a datos compartidos tiene cuatro elementos básicos, siendo;
    a. El derecho a seleccionar el contenido de los datos compartidos,
    b. El derecho a conocer las tecnologías,
    c. El derecho a acercarse a los canales de comunicación, y
    d. El derecho a la calidad del servicio garantizado.
2. Derecho a utilizar los datos compartidos: El propósito de este derecho no es que las personas tengan acceso a los datos, sino que los utilicen, pues a través del intercambio de datos, los poseedores de datos ayudan a otros a satisfacer sus necesidades y así obtener intereses debidos. Este derecho puede separarse de los derechos a compartir para convertirse en una función independiente y así servir como base importante para establecer

derechos de usufructuario de datos. Este derecho es una versión restringida del derecho a usar, pues el perseguir el valor del uso de datos y obtener el interés correspondiente a través de su uso, abarca principalmente el derecho a procesar datos compartidos y el derecho a reproducirlos. Es decir, el procesamiento es un método básico para utilizar los datos compartidos, por lo tanto, el derecho a procesar datos compartidos es una parte importante del derecho de utilizar datos compartidos.

3. *Jus fruendi* de los datos compartidos: Se refiere al derecho de las personas a obtener ganancias con los datos compartidos. Después de obtener acceso a los datos compartidos y utilizarlos, las personas también pueden obtener algunas ganancias de ellos. Sin embargo, el *jus fruendi* de los datos compartidos está sujeto a restricciones que se derivan de los objetos del derecho, es decir, los datos que se comparten, pues si bien el propósito de datos es obtener ganancias, el *jus fruendi* de los datos compartidos juega un papel importante en ampliar la escala y establecer el orden para transmisión de datos. Los datos compartidos solo pueden demostrar su valor cuando se obtiene el *jus fruendi* y esto prueba la externalidad y largo ciclo de retornos del *jus fruendi* de datos compartidos, así como la diversidad de partes involucradas. La externalidad proviene de la relación entre las ganancias obtenidas al usar, compartir los datos y costos pagados, por lo que deben imponerse restricciones necesarias al *jus fruendi* de los datos compartidos para lograr un equilibrio entre los intereses privados y públicos, pues si bien las ganancias de los datos compartidos tienen un ciclo largo, ya que el valor compartido de los datos no se puede explorar lo suficiente utilizado o consumiendo los datos una vez. Cuando se agregan nuevos recursos de datos y se conectan a lo que ya existe, el valor compartido de los datos aumentará y vendrán nuevas ganancias, empero, a la vez los datos compartidos no se desgastan, ya que pueden compartirse más y usarse una y otra vez, generando beneficios durante un largo período de tiempo, sin alcanzar nunca su punto máximo.

4. El derecho a la disposición de los datos compartidos: Esto representa el dominio final sobre los datos, siendo el derecho de las personas a que se eliminen los datos compartidos. Desde una perspectiva diferente, la disposición puede considerarse como el uso final de los datos, por lo tanto, este derecho es el derecho a consumir y compartir los datos. Esto no significa que quienes ofrecen los datos para compartir pierdan el control sobre los datos, más bien, es una parte independiente de los derechos de compartir formados a través de la reproducción de datos, que permite que varios titulares de los derechos de compartir tengan acceso y utilicen los mismos datos simultáneamente. Además, esto no elimina el valor original de los datos, al contrario, aumenta el valor compartido. Para los *res corporales*, la disposición conduce a la extinción absoluta o relativa de su propiedad y el derecho a la disposición de ellos determina su propiedad. En el caso de los datos, en cambio, el derecho tradicional a disponer ya no es aplicable. Va en contra de la tendencia de desarrollo de datos buscar poseer datos. Dado que el valor de los datos radica en que se comparten, el control de los datos no es el propósito de los derechos de compartir. Para resolver el problema de la escasez de datos y aprovechar al máximo el valor de los datos, es necesario ampliar la escala de los datos y permitir que más personas tengan la oportunidad de utilizarlos. El derecho a la disposición de los datos compartidos tiene más que ver con el uso de datos compartidos, que es la forma en que las personas persiguen el valor de los datos. Por lo tanto, el derecho a disponer de los datos compartidos es tan importante como la propiedad de los datos compartidos.

Volviendo al tema que nos ocupa, se debe hacer hincapié en que es cierto que la tecnología es importante, y que los postulados teóricos únicamente se quedan en postulados si no se cuentan con la suficiente tecnología actual para ejecutar de manera adecuada las ideas, empero, la tecnología por sí sola no es suficiente.

Por ejemplo, una solución tecnológica podría ser que un usuario de la nube pueda mantener la propiedad de sus datos a través de restricciones técnicas, como cifrar la información para que solo él o ella puedan usarla. Pero el cifrado tiene inconvenientes y ventajas, en particular que algunos aspectos del servicio en la nube pueden no ser utilizables con respecto a la información cifrada, asimismo, otra opción para los clientes sofisticados es utilizar una máquina virtual (VM) que se ejecuta en la infraestructura del proveedor de la nube, lo que permitirá al cliente, pero no al proveedor de la nube, acceder y utilizar la información (Reed, 2021, 199-201).

Otra solución con un enfoque más disruptivo es utilizar el *ledger* de *Blockchain* para compartir información (Treder, 2019, 31). Pero como se reiteró anteriormente, la tecnología por sí sola no es suficiente. Tal aseveración se respalda en el caso del año 2016, donde en la red *Blockchain Ethereum* un proyecto llamado *The DAO* fue hackeado, lo que provocó un debate sobre cómo responder. La comunidad decidió realizar un *hard fork* para revertir la transacción y recuperar los fondos robados, pero esto suscitó discusiones sobre la seguridad y la gobernanza en las redes *blockchain*, así como sobre los principios de inmutabilidad. El caso subrayó la importancia de la seguridad en las criptomonedas y planteó preguntas sobre la ética de cambiar la cadena de bloques para corregir errores (Werbach, 2018, 67-69), e indirectamente superponer la gobernanza administrativa sobre la gobernanza operativa.

En síntesis, independientemente de todos los postulados teóricos, es imperante salvaguardar los datos con derechos sustantivos, procesales, requiriendo el consentimiento, afrontando desafíos tecnológicos; conforme a la habilitación de la reutilización de datos públicos, con el establecimiento de políticas de datos abiertos, garantizando estándares de clasificación de datos unificados, permitiendo el acceso a la información y promoviendo la interoperabilidad de datos y sistemas; y en lo que respecta al sector privado, con la promoción de licencias abiertas, requiriendo portabilidad de datos, uso de API, forjar asociaciones de latos y limitando la responsabilidad de los intermediarios; finalmente como consideraciones finales, tendrán que abordarse las salvaguardas

de adoptar e implementar legislaciones, introducir modelos de consentimiento más significativos, ampliar la protección a datos mixtos y privacidad grupal, protección de datos por diseño y por defecto y priorizar las medidas de ciberseguridad (Banco Mundial, 2021, 193-230).

La principal razón por la que surgen conflictos entre el derecho a compartir y el derecho a la privacidad es por el conflicto entre los intereses públicos y los intereses personales y las discrepancias entre los intereses patrimoniales y los intereses de la personalidad, lo que crea un enorme problema en el ámbito del derecho y la legislación de derechos de datos (Yuming, 2021, 185-186). Una propuesta para abordar este conflicto es establecer un marco legal y ético que equilibre adecuadamente los intereses públicos y personales, así como los intereses patrimoniales y de la personalidad. Esto podría lograrse mediante la implementación de regulaciones claras y transparentes que definan notoriamente los derechos y responsabilidades de todas las partes involucradas. Además, se podría promover la educación y la conciencia pública sobre la importancia de la privacidad y la protección de datos, fomentando así un mayor respeto por estos derechos. Asimismo, se podría fomentar la innovación tecnológica para desarrollar soluciones que permitan compartir datos de manera segura y proteger la privacidad de los individuos al mismo tiempo. En última instancia, se requeriría un enfoque colaborativo entre gobiernos, empresas, sociedad civil y otros actores para abordar de manera efectiva este complejo problema en el ámbito del derecho y la legislación de derechos de datos.

Con miras hacia el futuro y siguiendo la teoría de derecho natural de Locke, todos deben tener derecho a la vida, libertad y propiedad que son los tres pilares de una sociedad moderna, pero con el surgimiento de los datos, se espera que los derechos a los datos se convierta en un cuarto derecho humano básico después de los derechos a la vida, propiedad y libertad, toda vez que el derecho a los datos no es un derecho simple, sino una colección de derechos que contienen los derechos de diferentes partes sobre el mismo objeto, y que involucra personalidad, privacidad, propiedad y soberanía (Yuming, 2021, 44-46).

1.5. Marco jurídico vigente para el acceso e intercambio de datos

Para abordar el marco jurídico vigente actual en materia de datos, es imperativo recurrir al máximo estándar, ya que siguiendo la premisa de "quien puede lo más puede lo menos", se garantiza un cumplimiento integral y se obtiene una representación precisa de las innovaciones y el consenso comunitario. Al adoptar el estándar más elevado, se establece una base sólida para la regulación de datos en un entorno cada vez más digitalizado y globalizado.

Cabe destacar que, en este análisis únicamente se abordarán las normativas de la UE, las cuales están liderando el camino hacia nuevas teorías en el ámbito de los datos. No obstante, se evitará adentrarse en detalle en estas regulaciones debido a que no forman parte del derecho mexicano. Además, es importante destacar que, con el tiempo, mencionar exclusivamente las normativas europeas vigentes podría resultar anacrónico, dado el constante cambio y desarrollo del marco jurídico global en materia de datos. Por ende, este enfoque será breve y se centrará en los aspectos más relevantes para el contexto actual.

Para empezar, el Consejo de Europa (1981) emitió el 28 de enero de 1981, el Convenio para la Protección de las Personas con Respecto al Tratamiento Automatizado de Datos de Carácter Personal, o conocido como Convenio 108. Mismo Convenio fue aceptado y ratificado por México, para ser promulgado el 28 de septiembre del 2018 (DOF, 2018).

Dicho Convenio 108, establece un marco para garantizar el respeto a los derechos y libertades fundamentales, especialmente el derecho a la privacidad, en el tratamiento automatizado de datos personales. Define términos clave como datos personales, ficheros y tratamiento automatizados, y establece disposiciones generales, principios básicos de protección de datos, reglas para categorías especiales de datos, seguridad de los datos, garantías para las personas concernidas, excepciones y restricciones, sanciones y recursos, entre otros. También aborda la cooperación entre los Estados parte, los flujos transfronterizos de datos, la asistencia mutua y la creación de un Comité Consultivo para facilitar la aplicación del Convenio. Cabe destacar que este Convenio se toma como el

primer instrumento internacional en materia de protección de datos personales. Cabe hacer la aclaración que dicho Convenio fue en el marco de Europa y no de la UE *per se*.

En el mismo orden de ideas, si bien la normatividad en protección de datos personales en la UE surgió en la Directiva 95/46/CE relativa a la protección de personas físicas en lo respectivo al tratamiento de datos personales y a la libre circulación de estos datos, actualmente se cuenta con una normatividad con carácter de Reglamento, es decir, de aplicación directa y obligatoria en todos los Estados miembros de la UE, siendo el **Reglamento General de Protección de Datos (RGPD) (2016)**, que tiene como objetivo principal proteger a las personas cuando sus datos son procesados por empresas, entidades públicas y privadas y personas que utilicen o gestionen cualquier dato personal en sus actividades profesionales. Permite a los individuos tener un mayor control sobre sus datos personales; moderniza y unifica las reglas, lo que beneficia a las empresas al reducir la burocracia y aumentar la confianza de las personas. Establece un sistema de autoridades supervisoras independientes para monitorear y hacer cumplir el cumplimiento. Este Reglamento surge para fortalecer los derechos existentes de los individuos, introducir nuevos derechos y otorgarles más control sobre sus datos personales, facilitando así el acceso a la información, la portabilidad de datos y el derecho al olvido. Además, implementa normas para las empresas, como la designación de un oficial de protección de datos y la simplificación de los procedimientos de notificación de brechas de seguridad, lo que crea un campo de juego nivelado y promueve la innovación.

Si bien el RGPD se destaca como el instrumento fundamental que regula la protección de datos personales, es importante destacar que su alcance se limita a los datos de carácter personal. Esto implica que, si bien brinda un marco robusto para proteger la privacidad de los individuos en el ámbito de la UE, no aborda otros tipos de información que pueden ser sensibles o relevantes en diferentes contextos, como datos empresariales, financieros o de propiedad intelectual. Por lo tanto, mientras que el RGPD establece una base sólida para la protección de la protección de datos en la era digital, también es esencial considerar otros marcos

regulatorios y mejores prácticas para abordar de manera integral la gestión de datos en todas sus formas.

Dicho lo anterior, ante tal ausencia, la misma UE promulgó el **Reglamento relativo a un marco para la libre circulación de datos no personales en la UE (2018)**, la cual establece un marco para garantizar el libre flujo de datos no personales en la UE, permitiendo el procesamiento libre de datos electrónicos, excluyendo los datos personales. Prohíbe restricciones sobre dónde pueden almacenarse o procesarse los datos, aplicándose a datos proporcionados como servicio a usuarios en la UE o procesados por individuos, empresas u organizaciones en la UE. Los países de la UE deben informar a la Comisión Europea de cualquier nueva restricción de localización de datos, y los instó a derogar restricciones injustificadas y establecer un punto de información nacional en línea sobre los requisitos de localización. Además, se anima a la creación de códigos de conducta auto-regulatorios a nivel de la UE para mejorar la cooperación y la conciencia sobre las mejores prácticas en el procesamiento de datos.

Cabe destacar que, de manera paralela la UE también ha promulgado directivas relacionadas a los datos. Por mencionar algunas está la Directiva relativa a los datos abiertos y la reutilización de la información del sector público (2019), o la Directiva sobre la privacidad y las comunicaciones electrónicas (2002), y que actualmente existe un proyecto para convertirla en Reglamento (Comisión Europea, 2017). Se debe aclarar que los reglamentos son vinculantes en su totalidad y tienen efecto inmediato, mientras que las directivas requieren que los Estados miembros las transpongan en su legislación nacional dentro de un plazo establecido.

Lo dicho hasta aquí presupone que la visión contemporánea de Europa y la UE, como actores prominentes en regulación, era concebir los datos desde el enfoque de la dignidad humana, basándose en la protección de datos, privacidad y datos abiertos. No obstante, fue hasta el año 2020 que la UE por medio de la Comisión Europea, emitió su propia Estrategia Europea de Datos, que buscaba crear un mercado único de datos en la UE para impulsar la competitividad y la

innovación. Esto implicaba facilitar la circulación de datos entre sectores, garantizando el cumplimiento de normativas de privacidad y competencia. La UE pretende establecer normas claras para el acceso y la reutilización de datos, invertir en infraestructuras tecnológicas, promover la colaboración en la computación en la nube y empoderar a los usuarios con herramientas de control sobre sus datos. Además, se propone un marco legislativo para la gobernanza de espacios de datos comunes y normativas específicas para conjuntos de datos de alto valor (Comisión Europea, 2020a). Ante esto, se podría inferir que es una transición ante una economía basada en datos, no tomando los datos sólo como datos personales, sino agregando datos sujetos a derechos de autor y secretos industriales.

Como bien se mencionó en la Estrategia de datos de la UE, la misión y visión de la UE iba acompañada de la propuesta de crear un marco legislativa para la gobernanza de los datos. En este sentido, posteriormente a la estrategia se promulgó el **Reglamento de Gobernanza de Datos** (2022), que tiene como objetivo principal proporcionar un marco para aumentar la confianza en el intercambio voluntario de datos en beneficio de las empresas y los ciudadanos. Reconoce el enorme potencial económico y social de los datos, especialmente en áreas como la salud, la innovación y la creación de empleo. Sin embargo, este potencial no se está realizando debido a obstáculos como la escasa confianza en el intercambio de datos y cuestiones técnicas. La ley aborda esto facilitando el intercambio de datos de manera confiable y segura. Se aplica tanto a datos personales como no personales, estableciendo normas para su reutilización, regulando los intermediarios de datos y fomentando el altruismo de datos. En la práctica, la ley establece requisitos técnicos para el sector público, promueve la neutralidad y transparencia de los intermediarios de datos, y permite que individuos y empresas compartan sus datos sin recompensa económica para el interés público. Además, crea el Comité Europeo de Innovación de Datos para facilitar el intercambio de mejores prácticas y establece salvaguardias para los flujos internacionales de datos.

Y, por otra parte, consecutivamente se expidió el **Reglamento de Datos** (2023), siendo una medida integral para abordar los desafíos y aprovechar las oportunidades de los datos, enfatizando el acceso justo y los derechos de los usuarios mientras garantiza la protección de los datos personales. Complementa el Reglamento de Gobernanza de Datos, regulando quién puede crear valor a partir de los datos y en qué condiciones, facilitando así el acceso fiable y seguro a los datos en sectores clave y ámbitos de interés público. Introduce medidas para aumentar la seguridad jurídica en el uso de datos, mitigar el abuso de desequilibrios contractuales, permitir a organismos públicos acceder a datos privados para fines de interés público, facilitar la interoperabilidad de datos y revisar aspectos de la Directiva sobre bases de datos. En la práctica, la Ley de Datos fomentará la innovación y la creación de empleo al garantizar el acceso a datos generados por dispositivos *IoT*, lo que permitirá mejoras en servicios posventa, eficiencia industrial y agricultura de precisión.

Resumiendo lo dicho hasta aquí, se podría inferir que hubo una transición de contemplar en las regulaciones los datos únicamente como datos personales y no personales (Teoría de un nuevo tipo de Derecho de Personalidad), para posteriormente crear una economía de los datos y considerar también datos sujetos a derechos de autor y secretos comerciales, para su reutilización y venta (Teoría de los derechos de Propiedad Intelectual; Teoría de los Secretos Comerciales; y Teoría de los Derechos de Propiedad de los Datos). Empero, tomando en cuenta las teorías legales de los datos mencionadas en el capítulo 1, faltarían dos concepciones y teorías legales de los datos, las cuales son la Teoría de los derechos de protección al consumidor; y Teoría de los derechos de competencia. Al respecto, la UE siguió su desarrollo y expidió dos Reglamentos que consideran los bienes jurídicos de estas dos últimas teorías en comento.

En primer lugar, se expidió el **Reglamento de Mercados Digitales** (2022), que tiene la finalidad de garantizar mercados digitales justos y abiertos. Dentro de sus múltiples funciones, las que destacan vinculadas a los datos son que las plataformas *gatekeeper* o guardián de acceso (dominan el tráfico y el negocio global en Internet) deben permitir a sus usuarios comerciales acceder a los datos

que generan en su uso de su plataforma. Además, se prohíbe la combinación de datos recopilados de dos servicios diferentes pertenecientes a la misma empresa, (por ejemplo, en el caso de Meta, sus productos Facebook y WhatsApp); también se establecen disposiciones para proteger a los usuarios comerciales de las plataformas (incluidos anunciantes y editores); se implementan instrumentos legales contra los métodos de auto preferencia utilizados por las plataformas para promocionar sus propios productos; se establecen disposiciones relacionadas con prácticas de agrupación; y se garantiza la interoperabilidad, portabilidad y acceso a datos para empresas y usuarios finales de las plataformas.

De manera paralela, vinculada a las teorías relacionadas a protección al consumidor y competencia económica, además del Reglamento de Mercados digitales, se expidió el **Reglamento de Servicios Digitales** (2022), que tiene como objetivo principal crear un entorno en línea más seguro para consumidores y empresas en la UE, mediante un conjunto de reglas diseñadas para proteger los derechos fundamentales de los consumidores, definir responsabilidades claras para las plataformas en línea y abordar contenido ilegal, discurso de odio y desinformación. Específicamente fortalece a los usuarios y la sociedad civil, incluyendo el acceso para autoridades e investigadores a datos clave generados por las plataformas muy grandes para evaluar riesgos en línea. También se aborda la evaluación y mitigación de riesgos, con medidas de protección para niños y límites en el uso de datos personales sensibles para publicidad dirigida.

Lo dicho hasta aquí demuestra que, hasta el momento, se ha presenciado una clara evolución en la UE, que se ha traducido en la expedición de regulaciones fundamentadas en diversas teorías legales. Desde la concepción de un nuevo tipo de Derecho de Personalidad, se ha transitado hacia la integración de la Teoría de los Derechos de Propiedad Intelectual (PI), la Teoría de los Secretos Comerciales, la Teoría de los Derechos de Propiedad de los Datos, la Teoría de los Derechos de Protección al Consumidor y la Teoría de los Derechos de Competencia. Este proceso ha dado lugar a la formación de una teoría ecléctica y complementaria sobre los derechos relacionados con los datos, que

refleja la complejidad y la interconexión de los diversos aspectos legales involucrados en la era digital.

Cabe destacar que otros países como EE. UU. han mezclado las teorías de un nuevo tipo de derecho personalidad y la Teoría de los derechos de protección al consumidor en una misma normatividad, con la *California Consumer Privacy Act*. En lo que respecta a México, se cuenta con una visión meramente vinculada a la teoría de un nuevo tipo de derecho personalidad, con la Ley Federal de Protección de Datos Personales en Posesión de los Particulares y su Reglamento (normativa sector privado); y la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (normativa sector público).

Para concluir, si bien pareciera que el debate sobre las teorías sobre los datos ha concluido y se perfeccionará de acuerdo al espectro de las normatividades actuales por parte de la UE, mientras que otras naciones como México podrán tomar su camino emitiendo regulaciones similares a la protección de bienes jurídicos, pero ajustados a la realidad nacional. Lo cierto es que, no se deben observar únicamente los datos de manera aislada, sino en complemento de diversas tecnologías de la industria 4.0, estando atentos a regulaciones específicas relativas a *blockchain*, sistemas *ciber-físicos*, impresión 3d, ciberseguridad, *big data*, realidad virtual y aumentada, computación en la nube, Internet de las Cosas (*IoT*), robots e Inteligencia Artificial (IA).

Dicho lo anterior, una regulación recién aprobada en el año 2024 es la **Ley de Inteligencia Artificial** (Comisión Europea, 2021). La cual no regula los datos *per se*, pero si los sistemas de IA que necesitan datos. La próxima Ley de Inteligencia Artificial establece normas y definiciones para el uso de sistemas de IA en diferentes contextos, especialmente enfocándose en la identificación biométrica remota. Se distingue entre sistemas en tiempo real y en diferido, con regulaciones específicas para cada caso, priorizando la protección de datos biométricos y derechos fundamentales. Los sistemas de IA de alto riesgo deben cumplir con requisitos rigurosos en cuanto a la calidad de los datos, la transparencia, la ciberseguridad y la mitigación de sesgos. Se promueve el acceso a conjuntos de datos de alta calidad para el desarrollo de sistemas de IA, asegurando prácticas

adecuadas de gobernanza y gestión de datos. Además, se detallan condiciones para el tratamiento de datos personales en entornos controlados de pruebas, así como disposiciones para el acceso y evaluación de sistemas de IA por parte de autoridades pertinentes.

1.6 Uso ético de los datos

El uso ético de los datos es un aspecto fundamental en el contexto de los avances tecnológicos, la proliferación del *Big Data* y las capacidades para recopilar, procesar y analizar grandes volúmenes de información. Este ámbito no se limita al cumplimiento normativo, sino que busca alinear las prácticas organizacionales con valores fundamentales que respeten los derechos de las personas y promuevan una innovación responsable. En este sentido, dos enfoques éticos predominan en la toma de decisiones relacionadas con los datos: el utilitarismo y la deontología. Según Martin (2022, 51-56), el utilitarismo prioriza decisiones que maximicen el bienestar colectivo, justificando, por ejemplo, el uso de sistemas de análisis de datos para optimizar procesos empresariales, incluso si ello perjudica a minorías. Por otro lado, la deontología se centra en principios universales como la dignidad humana y los derechos fundamentales, oponiéndose a prácticas que violen la privacidad o la autonomía individual, independientemente de sus beneficios. Ambos marcos resaltan la complejidad ética inherente a los contextos tecnológicos actuales.

En el contexto del *Big Data*, la capacidad tecnológica para recolectar datos masivos genera oportunidades importantes para la ciencia y la innovación, pero también plantea dilemas éticos. Collmann y Matei (2016, 1-7) señalan que esta capacidad frecuentemente deja las preocupaciones éticas como una reflexión posterior, poniendo en riesgo derechos fundamentales como la privacidad, la autonomía y la confianza pública en las instituciones científicas. Ante esta problemática, proponen un marco de razonamiento ético que permita anticipar y gestionar los dilemas antes de iniciar proyectos de investigación. Herramientas

como el *Privacy Matrix* resultan esenciales para identificar las implicaciones éticas y de privacidad en proyectos de gran escala, fomentando así la confianza pública.

El respeto por los derechos fundamentales y la promoción de la confianza pública son pilares esenciales para garantizar el manejo ético de los datos. Esto se logra mediante la aplicación de principios como la transparencia, la minimización de datos y el consentimiento informado. El RGPD establece principios fundamentales, como la limitación de propósito y la rendición de cuentas, que buscan proteger los derechos de los interesados y garantizar la responsabilidad de las organizaciones. Según Martens (2022, 19-40), la transparencia exige que las personas sean informadas de manera clara y comprensible sobre cómo se recopilan, procesan y almacenan sus datos, incluyendo detalles sobre el propósito del procesamiento y los posibles riesgos asociados. Sin embargo, lograr una transparencia efectiva puede ser un desafío debido a la complejidad de las políticas de privacidad, lo que Martens denomina la paradoja de la transparencia. La responsabilidad en este contexto implica que las organizaciones implementen medidas efectivas para garantizar el cumplimiento de los principios establecidos, lo que incluye auditorías regulares y la preparación para responder ante fallas en el sistema.

El consentimiento informado es otro pilar central en el manejo ético de los datos. Este debe ser obtenido de manera libre y con información objetiva sobre la naturaleza de la investigación, los riesgos potenciales y las alternativas disponibles. La falta de consentimiento informado puede resultar en violaciones éticas graves, como el uso de datos sensibles sin autorización para ajustar primas de seguros de salud, lo cual plantea considerables riesgos éticos y legales. Además, el principio de minimización de datos establece que solo se debe recopilar la información estrictamente necesaria para cumplir con fines específicos, garantizando que los riesgos asociados con la recopilación y el procesamiento sean reducidos al mínimo mientras se maximizan los beneficios potenciales.

El rápido avance tecnológico también ha generado desafíos éticos adicionales, según O'Keefe y O'Brien (2018, 259), quienes destacan que muchas

prácticas organizacionales se desarrollan sin una consideración adecuada de los derechos y expectativas de las personas afectadas. Para mitigar estos riesgos, proponen herramientas como los *Ethical Impact Assessments* (EIA), que permiten identificar y gestionar riesgos éticos a lo largo del ciclo de vida de los datos. En esta línea, los principios de *Privacy by Design*, subrayan la importancia de integrar la privacidad como configuración predeterminada en los sistemas tecnológicos, desde su diseño hasta su implementación. Este enfoque incluye el respeto por la privacidad del usuario y la promoción de la transparencia como un valor fundamental.

El uso ético de los datos no solo enfrenta retos en términos de diseño y aplicación tecnológica, sino también en el contexto del *Big Data*. La capacidad para procesar grandes volúmenes de información amplifica los impactos de las decisiones organizacionales, lo que incluye situaciones como decisiones automatizadas que afectan derechos fundamentales o el manejo erróneo de datos sensibles (Davis, 2012, 03). Para abordar estas problemáticas, se sugiere herramientas como los marcos de alineación de valores, que permiten a las organizaciones integrar principios éticos en sus operaciones diarias y minimizar incoherencias entre sus políticas internas y las percepciones externas. Además, las consultas éticas explícitas dentro de las decisiones operativas pueden prever y mitigar consecuencias no deseadas, fortaleciendo la confianza entre consumidores y socios comerciales (Davis, 2012, 06).

En conclusión, el uso ético de los datos es esencial para construir relaciones sostenibles y promover una innovación responsable. Las organizaciones deben reconocer que evitar discusiones éticas no elimina sus responsabilidades, sino que las amplifica, afectando la confianza de los consumidores y su reputación. Integrar prácticas éticas en el manejo de datos no solo responde a exigencias legales, sino que también refuerza los valores fundamentales que sostienen una sociedad justa y equitativa.

Capítulo 2

TIPOS DE LOS DATOS Y SU CLASIFICACIÓN LEGAL.

Capítulo 2. Tipos de datos y su clasificación legal

Tras haber examinado detalladamente las teorías sobre los datos y haber analizado el marco jurídico vigente en el capítulo previo, el siguiente paso consiste en explorar los distintos tipos de datos y su clasificación legal. Este enfoque permitirá profundizar en la comprensión de la diversa naturaleza de la información que se maneja en el contexto digital, así como en las normativas que regulan su tratamiento y protección. Al abordar estos aspectos, no solo se amplía el conocimiento sobre la gestión de datos, sino que también crea un mejor escenario para abordar los desafíos éticos, sociales y legales que surgen en la era de la información.

En la actualidad, resulta difícil obtener, utilizar y transmitir datos debido a la falta de claridad en sus conceptos y términos. A menudo, se trata a los datos como una entidad monolítica, a pesar de que la evidencia científica ha demostrado que los datos son un bien heterogéneo cuyo valor depende del contexto de uso, con distintas implicaciones para personas, empresas y gobiernos (OCDE, 2019a, 45). Además, con tecnologías disruptivas aumenta la complejidad de separar los datos personales de los no personales, debido a que las diferentes jurisdicciones adoptan diversas definiciones de datos personales y no personales, lo que causa superposiciones entre los tipos de datos (WEF, 2019c, 11).

Los procesos de datos enfrentan desafíos que involucran aspectos técnicos y legales. Aunque inicialmente estos requisitos pueden parecer independientes, están interconectados, lo que requiere una comunicación efectiva y un lenguaje común entre los interesados que usan datos. Desde el punto de vista legal, los datos se encuentran bajo diversos campos, como la protección de datos y la propiedad intelectual, entre otros. Estos campos a veces pueden entrar en conflicto, lo que lleva a reglas incompletas sobre el uso, reutilización, procesamiento y propiedad de datos. En consecuencia, identificar un camino rentable hacia el cumplimiento legal en el procesamiento de datos requiere tiempo y esfuerzo.

En el ámbito de la investigación de datos, existen varios marcos de clasificación para categorizar los datos, lo que proporciona bases valiosas para las investigaciones. Por ejemplo, la clasificación de Martin Abrams (2014, 6-9) incluye cuatro categorías: proporcionados, observados, derivados e inferidos. Además, Katharine Jarmul (2023, 4-5) ofrece un sistema de clasificación que abarca Origen de Datos/Linaje, Políticas y Controles, Confiabilidad/Conocimiento de Datos y Privacidad y Seguridad de Datos. Si bien estas clasificaciones son valiosas, es esencial reconocer que la clasificación de datos es un campo complejo y dinámico, y a veces es imposible proporcionar un ajuste simple para todos los tipos de datos con estos marcos.

Uno de los desafíos en la clasificación de datos surge de su inherente dinamismo. Los datos no son estáticos; evolucionan, se transforman y se adaptan con el tiempo. Esta naturaleza dinámica significa que los datos pueden atravesar diferentes clasificaciones a medida que cambia su fuente, contexto y propósito. Por ejemplo, datos que inicialmente son proporcionados pueden luego convertirse en derivados a medida que se extraen nuevos conocimientos de ellos mediante análisis. Este comportamiento dinámico difumina los límites de una simple categorización.

Otra consideración es que las clasificaciones de datos existentes pueden necesitar abordar de manera exhaustiva las necesidades de los tipos de datos emergentes y los casos de uso. A medida que el panorama de datos evoluciona, constantemente surgen nuevas formas de datos, cada una con sus características y requisitos únicos. Esta evolución desafía la completitud de las clasificaciones existentes. Por ejemplo, con la llegada de tecnologías como el Internet de las cosas (IoT) y la Inteligencia Artificial (IA), los datos pueden poseer atributos más allá de las categorías delineadas en marcos tradicionales. Además, los datos a menudo desafían la compartimentación estricta en categorías discretas. Los elementos de datos están frecuentemente interconectados, y sus atributos pueden superponerse a múltiples criterios de clasificación simultáneamente.

En resumen, aunque los marcos de clasificación de datos, como los introducidos por Abrams y Jarmul, ofrecen puntos de entrada valiosos para la

investigación de datos, es imperativo reconocer la esencia dinámica de los datos y el terreno constantemente evolutivo de los tipos de datos y sus aplicaciones. Los investigadores y profesionales deben abordar la clasificación de datos como un esfuerzo dúctil y receptivo, apreciando que los datos solo a veces se alinearán perfectamente con categorías predefinidas. Esta flexibilidad es fundamental para mantener el ritmo con el panorama de datos en constante cambio, garantizando que los esfuerzos de gestión e investigación de datos sigan siendo pragmáticos y pertinentes.

Se debe subrayar que no existen datos triviales y de ahí que estos no deben estar sujetos a la misma regulación. Por este motivo, la investigación futura sobre el derecho a los datos debe considerar la variedad de datos y el conjunto respectivo de derechos otorgados. Por ejemplo, en el caso de datos personales, existen datos sensibles que se encuentran particularmente cerca de la esfera personal, y por lo tanto, es posible que nunca sean asignables, mientras que otros datos pueden estar a tal distancia de la personalidad del sujeto que podría estar justificado la emancipación parcialmente del sujeto (Sattler, 2018, 47).

2.1. Teoría de Clasificación de Datos en Tres Niveles

Lo dicho hasta aquí presupone que además de que las teorías sobre los datos están incompletas, por su parte, la clasificación de datos no es exhaustiva. Por esta razón, la presente investigación, si bien no se busca establecer una taxonomía estricta de los datos, si pretende desarrollar una clasificación que tenga en cuenta diversos atributos, naturalezas y contextos de los datos, con el propósito de facilitar su gestión, gobernanza y la toma de decisiones pertinentes. Esta clasificación permitirá a las organizaciones comprender mejor la diversidad de los datos que manejan y cómo estos se relacionan con sus objetivos y requisitos específicos. Al considerar aspectos como la sensibilidad de los datos, su valor estratégico, el grado de confidencialidad requerido y el contexto operativo, se podrán implementar estrategias más efectivas para garantizar su adecuado

manejo y utilización, optimizando así los beneficios que pueden obtenerse de ellos.

El Reglamento de Gobernanza de Datos (2022) define datos como “toda representación digital de actos, hechos o información, así como su recopilación, incluso como grabación sonora, visual o audiovisual”, y de manera similar el Reglamento de Datos (2023) define igualmente datos como “cualquier representación digital de actos, hechos o información y cualquier compilación de tales actos, hechos o información, incluso en forma de grabación sonora, visual o audiovisual”.

Es cierto que las definiciones proporcionadas anteriormente aportan un marco inicial importante para comprender qué son los datos en términos generales. Sin embargo, la necesidad de una clasificación más detallada se vuelve imperativa en el contexto actual. Los datos pueden clasificarse de diversas formas según su naturaleza, su origen, su grado de sensibilidad o su utilidad.

No obstante, aunque sea necesario, en la práctica es complejo delimitar los datos en clasificaciones, toda vez que en algunas ocasiones ciertos datos podrían superponerse en diversas categorías y esto mermaría la clasificación, y por ende, se optó por tomar un modelo más flexible, que no limite los atributos de los datos, pero permita su identificación. Dicho lo anterior, se propone una Teoría de Datos respecto a su clasificación de Tres Niveles, a saber: **1. Obligatorio, 2. Buenas Prácticas y 3. Consideraciones Futuras**. Este marco tiene como objetivo categorizar los datos según su importancia legal, regulatoria y ética, orientando así a las organizaciones en sus procesos de manejo de datos.

En el primer nivel se encuentran los datos "**Obligatorios**", correspondientes a conjuntos de datos y actividades de procesamiento limitados por estrictos requisitos legales y regulatorios. El cumplimiento de estos mandatos no es opcional sino obligatorio y necesario para las organizaciones.

El segundo nivel se designa como "**Buenas Prácticas**", donde los datos caen en una categoría en la que el cumplimiento puede no ser estrictamente obligatorio, pero se considera altamente recomendable y alineado con las mejores prácticas.

El tercer y último nivel se refiere a las "**Consideraciones Futuras**", que abarcan datos y actividades de procesamiento que están emergiendo de una manera potencialmente significativa desde el punto de vista legal y ético.

En síntesis, la Teoría de Datos de Tres Niveles ofrece un enfoque estructurado para la gestión de datos y el cumplimiento normativo, permitiendo a las organizaciones categorizar sus datos según el nivel de escrutinio legal y regulatorio que requieren, desde los requisitos obligatorios hasta las buenas prácticas y consideraciones futuras. Al adoptar este marco, las organizaciones pueden manejar mejor el difícil panorama de la gobernanza de datos y mantener una ventaja competitiva en un mundo cada vez más impulsado por los datos.

Figura 3. Teoría de clasificación de datos de 3 niveles.

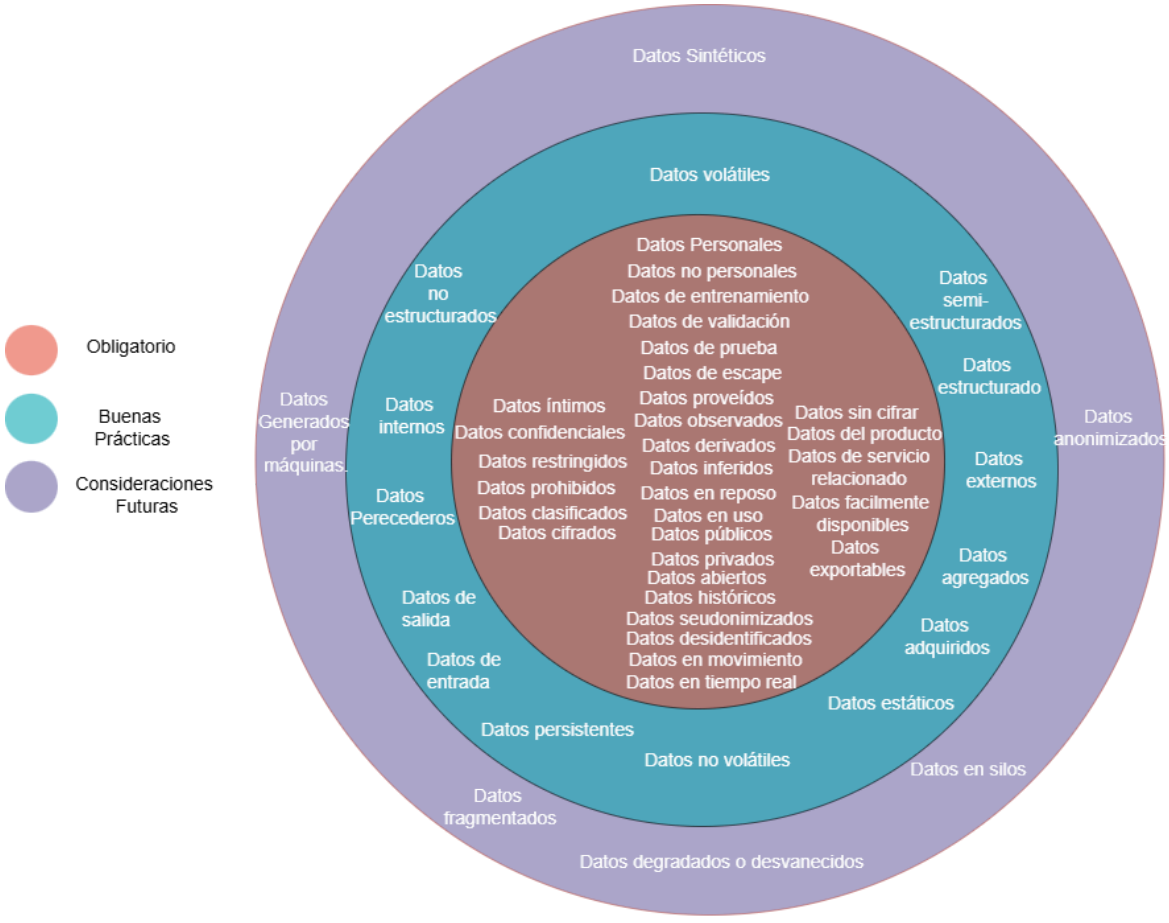

Figura de elaboración propia.

Este marco teórico introduce un enfoque estructurado para entender y organizar datos dentro del inmenso ámbito de la gobernanza de la información. Compuesto por tres niveles distintos, a saber, Obligatorio, Buena Práctica y Consideraciones Futuras, este marco busca revolucionar cómo se perciben y priorizan los datos. Este constructo teórico está diseñado para proporcionar claridad y dirección en una era definida por la toma de decisiones impulsada por datos y el aumento del escrutinio de las prácticas de manejo de la información. La Teoría de Datos en Tres Niveles capacita a las organizaciones para manejar el complejo panorama de gestión de datos con precisión y propósito al categorizar los datos según su importancia intrínseca.

2.1.1 Nivel Obligatorio

En el complejo escenario de la gestión de datos, existe un nivel que demanda atención y cumplimiento inquebrantable: el Nivel Obligatorio. Dentro de este ámbito, se encuentran conjuntos de datos y actividades de procesamiento regidos por rigurosas estipulaciones legales y regulatorias. El cumplimiento de estos mandatos trasciende el ámbito de la elección; es una obligación convincente, un imperativo vital y una necesidad irrefutable para organizaciones de toda índole. Estos requisitos tienen su origen en las leyes, regulaciones y normas de la industria, que dictan principios y reglas precisas de manejo de datos.

Entre estas normativas obligatorias, se encuentran regulaciones de protección de datos, verbigracia, el RGPD y la Ley de Portabilidad y Responsabilidad del Seguro de Salud (HIPAA).

Este categoría de nivel obligatorio, incluye datos personales; datos no personales; datos de entrenamiento; datos de validación; datos de prueba; datos de escape y *metadata*; datos proveídos; datos observados; datos derivados; datos inferidos; datos adquiridos; datos seudonimizados; datos desidentificados; datos agregados; datos anonimizados; datos en movimiento; datos en tiempo real; en reposo; en uso; datos públicos; datos privados; datos abiertos; datos históricos; datos íntimos; datos confidenciales; datos restringidos; datos prohibidos; datos

clasificados; datos encriptados; datos no encriptados; datos del producto; de servicios relacionados; fácilmente disponibles; y exportables.

2.1.1.1 Datos personales y no personales

Con respecto al primer tipo de datos, es necesario comenzar con los **datos personales** y en consecuencia, desentrañar la definición de datos personales, que consta de cuatro elementos: "toda información", "sobre", "identificada o identificable" y "persona física" (WP29, 2007, 6). Estos componentes ofrecen una interpretación amplia de los datos personales, abordando su alcance y las condiciones para su identificabilidad, mientras enfatizan que se refiere a individuos vivos.

Se requieren más detalles para describir adecuadamente los cuatro componentes de la definición de datos personales mencionar anteriormente. El primer componente, "toda información", abarca tres aspectos: naturaleza, contenido y formato. Incluye cualquier tipo de declaración sobre una persona, ya sea factual u basada en opiniones, incluso si no es necesariamente cierta, abarcando información sobre la vida privada de un individuo y diversas actividades. El segundo componente, "sobre", puede entenderse a través del contenido, propósito y resultado. El contenido se refiere a la información real, como los datos personales en un documento, el propósito involucra por qué se está procesando la información, como llevar a cabo una prueba de salud, y el resultado se refiere a cómo el procesamiento del contenido afecta los derechos de un individuo. El tercer componente, "identificada o identificable", significa que un individuo puede ser identificado directamente por su nombre o indirectamente a través de otra información. Por ejemplo, tener solo una foto puede no identificar a alguien, pero combinarla con otros datos de diversas fuentes puede distinguirlos. Por último, el componente “persona física” se refiere a personas naturales vivas, no a personas fallecidas, gobiernos o empresas (WP29, 2007, 6-26).

Es importante señalar que, sin perjuicio de los derechos de reputación y honor, existen teorías que sostienen que el derecho a los datos personales no

debería perderse al fallecer, especialmente en sistemas legales que consideran los datos personales como un derecho fundamental y no solo como un enfoque de consumo (Montes, 2020, 35-50). Sin embargo, independientemente de este debate teórico, tomando el RGPD como referencia para este trabajo de investigación, los datos personales solo serán considerados para personas vivas, señalando que puede haber excepciones en los marcos legales de algunas jurisdicciones.

Cabe destacar que, el concepto de identificación no solo implica la posibilidad de recuperar el nombre o la dirección de una persona, sino que también incluye la identificabilidad potencial mediante la individualización, la vinculación o la inferencia (WP29, 2014, 12). Los principales riesgos de la identificabilidad incluyen la capacidad de individualizar a una persona dentro de un conjunto de datos, vincular información o registros en una base de datos y hacer inferencias que puedan identificar a alguien basándose en valores o atributos. Por lo tanto, a simple vista y en la práctica, se pudiera considerar que no se trata un dato personal, pero si se puede individualizar, vincular o inferir la información de una persona física, puede que sea considerado dato personal y se deba cumplir con la ley en la materia.

En síntesis, los datos personales identifican a un individuo y se pueden usar para autenticación o autorización, mientras que los **datos no personales** son información no relacionada con detalles personales. Existen regulaciones para los datos no personales, como el Reglamento de la UE 2018/1807 para la libre circulación de datos no personales. Los datos no personales pueden convertirse en datos personales cuando se vinculan a una persona identificable, a menudo denominada "identificabilidad".

Por ejemplo, un experto debe tratar los datos personales, como nombres o historiales médicos, con estrictas medidas de seguridad, como cifrado, alguna base de tratamiento como el consentimiento explícito y acceso limitado, para proteger la privacidad y cumplir con normativas como el RGPD. Por otro lado, los datos no personales, como estadísticas anonimizadas, pueden gestionarse de manera más flexible, fomentando su reutilización para investigación o innovación,

asegurándose de que no puedan volverse identificables al combinarse con otros datos.

2.1.1.2 Datos de entrenamiento, validación y de prueba

La ley en IA (2024), define que los **datos de entrenamiento** son los datos usados para entrenar un sistema de IA mediante el ajuste de sus parámetros entrenables, entre los que se incluyen los pesos de una red neuronal. Al respecto, los datos de entrenamiento en sí no suelen ser datos personales, pero las organizaciones y los profesionales de datos deben ejercer precaución y adherirse a los principios de privacidad de datos y ética durante todo el ciclo de vida de los datos para mitigar el riesgo de manejar inadvertidamente datos personales y proteger los derechos de privacidad de las personas.

Respecto al segundo término, los **datos de validación** son datos usados para proporcionar una evaluación del sistema de IA entrenado y adaptar sus parámetros no entrenables y su proceso de aprendizaje, entre otras cosas, para evitar el sobreajuste. El conjunto de datos de validación puede ser un conjunto de datos independiente o formar parte del conjunto de datos de entrenamiento, ya sea como una división fija o variable (Comisión Europea, 2021). En el mismo sentido, los datos de validación no son inherentemente datos personales, pero los profesionales y las organizaciones deben ejercer precaución, seguir las mejores prácticas de privacidad de datos y considerar el potencial de revelación indirecta de información sensible al manipular y utilizar datos de validación en el contexto del aprendizaje automático y el análisis de datos.

Con relación al tercer concepto, los **datos de prueba** son datos usados para proporcionar una evaluación independiente del sistema de IA entrenado y validado, con el fin de confirmar el funcionamiento previsto de dicho sistema antes de su introducción en el mercado o su puesta en servicio (Comisión Europea, 2021). Los conjuntos de pruebas pueden contener datos personales, los cuales pueden ser utilizados para inferencias, toma de decisiones y mejora de modelos de inteligencia artificial. Las pruebas son cruciales en el desarrollo de sistemas de

IA, ayudando a seleccionar los componentes adecuados para cumplir con los requisitos de diseño y desarrollo (AEPD, 2020c, 36-37).

Por ejemplo, los datos de entrenamiento, aunque generalmente no son personales, requieren tratarse con precaución para evitar riesgos de privacidad y deben ser gestionados con principios legales durante todo su ciclo de vida. Los datos de validación, utilizados para evaluar y optimizar modelos, deben protegerse contra posibles revelaciones indirectas de información sensible, siguiendo las mejores prácticas de privacidad. Finalmente, los datos de prueba, que pueden contener información personal, requieren un manejo cauteloso, asegurando la anonimización cuando sea posible y el cumplimiento normativo, especialmente si son utilizados para inferencias o toma de decisiones

2.1.1.3 Datos de escape y *metadata*

Por otra parte, los **datos de escape** o *exhaust data,* son aquellos que son un subproducto de un proceso cuyo propósito principal es otro, por ejemplo, por cada imagen compartida, tuiteada, retuiteada o marcada como "me gusta", se genera una variedad de datos de escape: la persona que lo compartió, lo vio, el dispositivo utilizado, la hora, entre otros. Los **metadatos** son uno de los tipos más comunes de datos de exhaustividad porque describen las estructuras y propiedades de otros datos (Kelleher & Tierney, 2018, 242-243). De hecho, el Reglamento de Datos (2023), define los metadatos como “una descripción estructurada del contenido o de la utilización de los datos que facilita la búsqueda o la utilización de esos datos”.

Estos datos rastrean las actividades de los usuarios en Internet u otro sistema automatizado sobre comportamiento y transacciones. Sin duda, esto puede comprometer la privacidad, ya que ofrece información esencial sobre los hábitos del usuario. Por sí mismo, los datos de escape no son datos personales porque no identifican a individuos directamente, sin embargo, estos datos pueden convertirse en datos personales cuando se combinan con otra información, lo que permite la identificación de una persona física.

Aunque estos datos no suelen ser personales por sí mismos, su combinación con otras fuentes puede permitir la identificación de personas. Por lo tanto, deben gestionarse bajo principios de minimización y anonimización, asegurando que se limite su uso a lo estrictamente necesario y protegiendo contra la reidentificación. Además, su almacenamiento y procesamiento deben cumplir con las normativas aplicables, para garantizar un uso ético y respetuoso de la privacidad de los usuarios, especialmente en contextos de análisis de comportamiento o transacciones digitales.

2.1.1.4 Datos proveídos, observados, derivados e inferidos

Los **datos proveídos** se originan a través de acciones directas tomadas por el individuo conscientemente de las acciones que llevaron al origen de los datos, con tres categorías: 1. Iniciados, cuando las personas realizan una acción que inicia una relación, ya sea laboral, comercial, electoral o administrativa; 2. Transaccionales, cuando una persona está involucrada en una transacción, como pagar una factura, responder una pregunta o realizar una prueba; y 3. Publicados cuando las personas se expresan de manera proactiva y son conscientes de que están creando expresiones que serán vistas o escuchadas por todos, ampliando este espectro a través de las redes sociales (Abrams 2014, 6-9).

Los **datos observados** se observan y registran; por ejemplo, en Internet, se puede observar de dónde es una persona y el contenido revisado. Además, con Internet de las cosas (*IoT*), es posible observar el mundo físico y su conexión con el mundo digital. Tiene tres subcategorías. 1. Comprometidos, incluyendo datos como cookies en línea, donde el individuo se da cuenta de la observación en algún momento; 2. No Anticipados, siendo instancias en las que las personas son conscientes de que existen sensores pero tienen poco sentido de que los sensores estén creando datos de sus acciones; 3. Pasivos son aquellos creados inconscientemente, por ejemplo, cámaras en lugares públicos cuando se combinan con reconocimiento facial, donde la persona apenas es consciente de estar siendo observada (Abrams 2014, 6-9).

Los **datos derivados** resultan del procesamiento de los datos consigo mismos o uniendo estos con otras fuentes de información. Por ejemplo, podría resultar de operaciones aritméticas como el total gastado en una tienda o agrupando a un usuario en base a atributos compartidos en una base de datos. Por otro lado, los **datos inferidos** son el producto de aplicar un modelo a los datos, por ejemplo, puntajes de riesgo crediticio o de fraude (Abrams 2014, 6-9).

En ocasiones, se pueden tener datos no personales *ex-ante*, pero una vez que se procesan, el resultado puede convertirse en información personal *ex-post*. Es importante destacar que según el Artículo 20 del RGPD, solo los datos proporcionados y observados son elegibles para el derecho a la portabilidad de datos; esto excluye los datos inferidos y derivados. El sujeto de datos puede ejercer su derecho a la portabilidad de datos basándose en el consentimiento y el contrato como fundamentos para el procesamiento.

Finalmente, previo a dar por terminado el análisis de la teoría de Abrams, es importante señalar un tipo de datos que no está contemplado de manera estricta, siendo los **datos adquiridos**. Si bien podría hacerse una interpretación extensiva y decir que son parte de los datos proveídos, lo cierto es que, en los datos proveídos se necesita un acto comisivo del interesado de los datos, mientras que en los datos adquiridos es un acto comisivo delegado.

En este sentido, los datos adquiridos son información generada, obtenida, comprada o licenciada de terceros mediante contratos comerciales (por ejemplo, adquiridos de intermediarios de datos) u otros medios no comerciales (por ejemplo, adquiridos a través de iniciativas gubernamentales abiertas). Como resultado, las obligaciones contractuales y legales pueden afectar la reutilización y el intercambio de datos (OCDE, 2019a, 31). En ocasiones, sin que los usuarios se den cuenta, las aplicaciones venden regularmente la información de ubicación de los usuarios a otras empresas de terceros, que la utilizan con fines de *marketing* y otros propósitos; estas empresas de terceros recopilan e intercambian los datos con agencias gubernamentales (ACLU, 2022). Por esa razón, algunas regulaciones, como la Ley de Privacidad del Consumidor de California o en inglés *California Consumer Privacy Act* (2018), establecen que las empresas que venden

información personal están sujetas al requisito de la CCPA de proporcionar un enlace claro y conspicuo de "No Vender o Compartir Mi Información Personal" en su sitio web que le permite enviar una solicitud de exclusión, y las empresas no pueden exigirle que cree una cuenta para enviar su solicitud.

Continuando con la investigación y dejando a un lado la teoría de Abrams, ahora, es pertinente estudiar tipos de datos vinculados técnicas utilizadas para procesar datos y proteger la privacidad de los individuos. Estos datos son: datos seudonimizados, datos anonimizados, datos desidentificados, y datos agregados.

Para comenzar, de acuerdo con el artículo 4 del RGPD, la seudonimización es el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable.

Por ejemplo, los datos proveídos requieren consentimiento y gestión transparente, ya que son proporcionados directamente por los usuarios. Los datos observados, como los generados por sensores o cámaras, exigen medidas claras de notificación y anonimización, especialmente en casos de observación pasiva. Para los datos derivados e inferidos, que resultan del procesamiento y análisis, es fundamental garantizar que los resultados no comprometan la privacidad y que su uso respete los derechos de los usuarios, particularmente en contextos donde se pueden aplicar decisiones automatizadas sin intervención humana valorativa.

2.1.1.5 Datos seudonimizados, anonimizados, desidentificados y agregados

Los **datos seudonimizados** no pueden atribuirse a una persona específica sin el uso de información adicional (WEF, 2021b), y aunque la aplicación de la seudonimización a los datos personales puede reducir los riesgos para los sujetos de datos involucrados y ayudar a los controladores y procesadores a cumplir con sus obligaciones de protección de datos (ENISA, 2021), cabe hacer la aclaración,

que de facto este tipo de datos si cuentan con identificadores indirectos, y por lo tanto, si son considerados como datos personales.

Por otra parte, se encuentran los **datos anonimizados**. Una forma de describir datos anonimizados es compararlos con datos seudonimizados. La principal diferencia radica en la aplicabilidad de la ley a ellos. Si los datos seudonimizados ya no pueden atribuirse a un sujeto de datos específico (a menos que se utilice información adicional), el sujeto de datos sigue siendo indirectamente identificable, siendo datos personales y, por lo tanto, sujetos a la ley. Por el contrario, si se eliminan todos los elementos de identificación, se anonimiza y la ley no es aplicable. Sin embargo, en la práctica, distinguir entre estos dos tipos de datos es difícil, especialmente cuando muchos servicios o tecnologías utilizan el término "anonimizado" cuando en realidad son "seudonimizados" (Jasmontaitė-zaniewicz et al., 2021, 36).

El considerando (26) del RGPD establece que los principios de protección de datos no deben aplicarse a la información anónima, es decir, información que no se refiere a una persona natural identificada o identificable o datos personales anonimizados de tal manera que el sujeto de datos ya no sea identificable. Existen prácticas y técnicas de anonimización con diversos grados de robustez; algunos ejemplos son la Aleatorización (Adición de ruido, Permutación, Privacidad diferencial) y la Generalización (Agregación y anonimato k, Diversidad l, proximidad t) (WP29, 2014, 12-21).

En cambio, los **datos desidentificados** o despersonalizados significan datos de los cuales se han eliminado los valores de atributos que permiten la identificación directa y se han reemplazado por un identificador único para ocultar la identidad del sujeto de datos (Datatilsynet, 2015). Asimismo, según el NIST (s.f.), la información *desidentificada* son registros que han tenido suficiente información de identificación personal (PII) eliminada u ocultada de manera que la información restante no identifica a un individuo y no hay una base razonable para creer que la información pueda ser utilizada para identificar a un individuo.

A *prima facie*, los términos "desidentificación" y "anonimización" a veces se usan indistintamente, pero "anonimización" puede referirse a un tipo específico de

desidentificación que no puede revertirse. En el ámbito de la salud, "desidentificación" y "seudonimización" a menudo se consideran equivalentes, mientras que "anonimización" implica la eliminación del vínculo entre seudónimos e identidades de los sujetos (Garfinkel, 2015, 2-4). Mientras qué, en Noruega, especialmente en el sector de la salud, en la Ley de Sistemas de Archivo de Datos de Salud Personal, las definiciones de datos de salud seudonimizados y desidentificados fueron reemplazadas por el término más amplio datos de salud indirectamente identificables (Datatilsynet, 2015).

El concepto de datos desidentificados es algo complejo de definir porque a veces se utiliza como sinónimo de datos anónimos (Parlamento Europeo, 2020), y en ocasiones, el concepto de datos desidentificados se usa como sinónimo de datos anonimizados o seudonimizados, pero sus definiciones son diferentes. Por ejemplo, la anonimización es el acto de hacer que los datos personales sean anónimos; la seudonimización es la sustitución de parámetros directamente identificables por seudónimos, que aún constituirán indicadores de identificación únicos; y la desidentificación es la eliminación de todas las características personalmente identificables de los datos para que ya no se puedan vincular a un individuo específico (Datatilsynet, 2015). Por lo tanto, esto requiere una evaluación del riesgo de identificabilidad (o re-identificabilidad), un concepto en constante evolución (WEF, 2018).

A consideración autoral, la desidentificación es un proceso reversible, mientras que la anonimización es irreversible. Un ejemplo de esto es cuando investigadores de la Universidad de Austin, Texas, demostraron una nueva clase de ataques a datos estadísticos aplicados al conjunto de datos del Premio Netflix de 2006, que contenía calificaciones de películas *desidentificadas* de 500,000 suscriptores de Netflix (Narayanan & Shmatikov, 2008). El artículo y el caso demostraron cómo alguien con información limitada sobre un suscriptor individual podría identificar fácilmente a esa persona en un conjunto de datos. Utilizaron bases de datos públicas de películas en Internet, como IMDB, para recopilar conocimientos básicos, revelando información sensible como preferencias políticas y usuarios conocidos.

En síntesis, la desidentificación implica eliminar identificadores personales de los datos, lo que hace posible la reidentificación con fuentes adicionales. Por otro lado, la anonimización corta todos los vínculos entre una persona y los datos, lo que hace imposible la reidentificación. La desidentificación deja margen para la reidentificación, mientras que la anonimización asegura que no haya marcha atrás.

Finalmente, los **datos agregados** son información estadística sobre varios individuos que se han combinado para mostrar tendencias generales o valores sin identificar a los individuos dentro de los datos (ICO, 2012). Además, para aplicaciones de big data e inteligencia artificial, se utilizan datos agregados para comprender patrones en la sociedad, automatizar procesos y tomar decisiones (FRA, 2019).

La agregación de datos implica combinar conjuntos de datos individuales para analizar tendencias, preservando la privacidad al agrupar individuos con características similares. Para prevenir la reidentificación, un conjunto de datos agregados efectivo debe tener una población grande, categorizaciones amplias y excluir datos individuales únicos (IAPP, s.f.). La agregación es un proceso estadístico común en el que la información se agrupa y resume, impidiendo el acceso individual. Sin embargo, aún puede revelar detalles personales a través del análisis y filtros específicos, lo que potencialmente puede llevar a la identificación no intencionada. Aunque la agregación tiene como objetivo proteger la privacidad, no siempre garantiza un anonimato completo (Mcintosh, 2019).

En resumen, las diferencias entre los datos seudonimizados, anonimizados, desidentificados y agregados se representan en la siguiente tabla (Kelsey, 2016):

Cuadro 1. Diferencias entre datos seudonimizados, anonimizados, anonimizados y agregados.

Dato	Características	Ejemplo	Acción (identificadores)	Diferencias

Anonimización (no datos personales)	La eliminación o manipulación de identificadores directos e indirectos, así como de salvaguardias matemáticas y técnicas, se ha llevado a cabo para evitar la reidentificación.	Por ejemplo, se estima el ruido en un conjunto de datos para ocultar si un individuo está presente (privacidad diferencial).	Eliminación.	Irreversible
seudonimización (datos personales)	Información de la que se han eliminado o transformado los identificadores directos, pero los identificadores indirectos permanecen intactos.	Los seudónimos artificiales únicos reemplazan a los identificadores directos (por ejemplo, conjuntos de datos limitados de HIPAA, John Doe = 5L7T LX619Z) (secuencia única que no se usa en ningún otro lugar).	Sustitución	Posible reidentificación
Desidentificación (no datos personales)	Se han eliminado o manipulado identificadores indirectos, directos y conocidos para romper el vínculo con	Los nombres se eliminan de la base de datos de calificaciones de Netflix; sin embargo, hay información de otras	Separación	Posible reidentificación

	identidades del mundo real.	fuentes, como IMDb.		
Agregación. (no datos personales)	Se han combinado datos estadísticos sobre varios individuos para mostrar tendencias o valores generales sin identificar a los individuos dentro de los datos.	Se combinan diversos datos públicos y privados para producir estadísticas, como el número de pasajeros del transporte.	Combinación	Posible reidentificación

Cuadro de elaboración propia.

Por ejemplo, los datos seudonimizados, que contienen identificadores indirectos, deben ser protegidos como datos personales, asegurando que la información adicional necesaria para reidentificar a las personas esté bajo estrictas medidas de seguridad. Los datos anonimizados, al ser irreversibles, no se consideran personales, pero es esencial garantizar técnicas robustas de anonimización para prevenir la reidentificación. Los datos desidentificados requieren controles rigurosos, ya que su naturaleza reversible los hace vulnerables a la reidentificación con fuentes adicionales. Finalmente, los datos agregados, utilizados para análisis estadísticos, deben configurarse con poblaciones suficientemente grandes y categorizaciones amplias para evitar la identificación no intencionada.

2.1.1.6 Datos en movimiento, tiempo real, tránsito, reposo y uso

Pasando a otro tema, el estado de los datos es diverso; los datos pueden cambiar de estado rápidamente y con frecuencia, o permanecer en un solo estado a lo

largo de su ciclo de vida (Fitzgibbons, 2019). Por lo tanto, comprender sus características y diferencias entre los estados de los datos puede ayudar a las entidades a gestionar la información de manera más adecuada (Fitzgibbons, 2019). En este sentido, ahora es pertinente abordar los datos en movimiento, en tránsito, en reposo, y en uso.

Primero, los **datos en movimiento** son aquellos que se están trasladando a través de una red o en la memoria para su procesamiento en tiempo real (Hurwitz, et al., 2016, 26). Cabe destacar que, las nuevas fuentes de datos se están creando y añadiendo a la red todos los días. Por lo tanto, su valor principal radica en ser analizado poco después de ser creado, en muchos casos, inmediatamente después de ser creado; por lo tanto, el paradigma tradicional de gestión de datos donde los datos sin procesar se almacenan primero y se analizan después no se ajusta a la naturaleza temporal de los datos en movimiento (Rayes & Salam, 2019, 172-173).

Otro concepto similar es el de **datos en tiempo real**, a menudo referido como RTD, que son datos que se actualizan según su programación (por ejemplo, cotizaciones de acciones, estadísticas de fabricación, carga de servidores web y actividad de almacén) (Microsoft, 2014), y circulación de datos, que tiene tres formatos principales: puesta en común de datos, intercambio de datos y transacción de datos.

Continuando con los datos en movimiento, un punto a abordar son los **datos en tránsito**, ya que, a primera vista, esto podría ser sinónimo de datos en tiempo real o datos en movimiento; sin embargo, tiene un detalle que lo diferencia de este último. Los datos en tránsito se refieren a la información transferida a través del País A en el curso de su transporte hacia el País B o territorio fuera del País A sin que la información o datos personales sean accedidos, utilizados o revelados por ninguna organización (que no sea la organización que transfiere o un empleado de la organización que transfiere actuando en el curso de su empleo con la organización que transfiere) mientras los datos personales estén en el País A, excepto para tal transporte. Un ejemplo de datos en tránsito sería datos del extranjero que pasan por servidores dentro del País A hacia un destino en el

extranjero. Una organización que transfiere datos personales al extranjero se considerará que cumple con la obligación de limitación de transferencia con respecto a los datos en tránsito (PDPC, 2021). Por lo tanto, la transferencia no es lo mismo que el simple tránsito; es el procesamiento en el tercer país lo que completa la transferencia (Ustaran, 2019, 135). Como recomendación, se aconseja llevar a cabo una Evaluación de Impacto de la Transferencia (TIA, por sus siglas en inglés) antes de transmitir o transferir los datos (personales o no).

Por otra parte, los **datos en reposo** son aquellos que se almacenan en lugar de utilizarse en tiempo real (Hurwitz, et. al., 2016, 283). También se les conoce como datos almacenados (Moschovitis, 2021, 47). Además, los datos en reposo se refieren a los datos almacenados en sistemas estables y comprende datos recopilados y almacenados en estructuras como hojas de cálculo, bases de datos, almacenes, archivos y copias de seguridad, y datos móviles. Puede haber diferentes puntos donde se analizan los datos y donde se toman medidas (Arora, 2018, 16). Incluso, en algunas organizaciones pueden incluir un equipo de arquitectos que consideren los tres aspectos clave de la protección de datos: seguridad, privacidad y almacenamiento de datos (De Guise, 2020, 36). Por lo tanto, cuando los datos están en reposo, es necesario aplicar controles de seguridad adecuados para garantizar la confidencialidad y la integridad continua de la información, por ejemplo, utilizando cifrado (Gordon, 2016, 36).

En la práctica, los conjuntos de datos a menudo mezclan datos personales y no personales debido a recursos limitados, lo que lleva a que las pequeñas empresas traten todos los datos con altos estándares de cumplimiento. Esto resulta en que las leyes destinadas a los datos personales se apliquen a todos los tipos de datos (WEF, 2020a). Se debe subrayar que, el almacenamiento de datos, incluso si no se utiliza, se considera procesamiento de datos personales. Los datos personales deben tener periodos de retención estándar y deben revisarse, eliminarse o anonimizarse regularmente cuando ya no sean necesarios, pero la retención prolongada puede aplicarse por razones de interés público, investigación o fines históricos (ICO, 2021).

Crear una política de almacenamiento de datos que incluya un calendario de retención identificando diversos tipos de registros y sus periodos de retención requeridos, se materializa en un cumplimiento legal adecuado (Overly, 2021, 294-300). Por ejemplo, en el Estándar de Seguridad de Datos de la Industria de Tarjetas de Pago (PCI DSS) se contempla la protección de los Datos del titular de la Tarjeta, incluyendo la protección de datos en reposo y en tránsito (Viegas, et al., 2022, 39-42).

Por otra parte, los **datos en uso** se refieren a los datos en la memoria o los búferes de almacenamiento temporales; mientras una aplicación los está utilizando, no puede procesar datos encriptados; por lo tanto, debe descifrarlos en la memoria, de la mano de los controles de autenticación y autorización sólidos que ayuden a prevenir el acceso no autorizado (Chapple et al., 2018, 168).

Al respecto, desde una perspectiva de computación en la nube, un usuario puede establecer restricciones técnicas como la encriptación de información para que solo una persona autorizada pueda utilizarla. Sin embargo, la encriptación tiene sus ventajas y desventajas, especialmente porque algunos aspectos del servicio en la nube pueden volverse inutilizables en relación con la información encriptada; por lo tanto, otra opción más sofisticada es utilizar una máquina virtual ("VM") que se ejecute en la infraestructura del proveedor de la nube, permitiendo al cliente, pero no al proveedor de la nube, acceder y utilizar la información (Reed, 2021, 178), o por otra parte aplicar la cifrado homomórfico, donde los datos se encriptan antes de compartirse de manera que puedan ser analizados pero no descifrados en la información original (WEF, 2019d).

Prosiguiendo el análisis de datos obligatorios o vinculantes, surge la interrogante respecto a ¿Los usuarios confían en la precisión y confiabilidad de sus datos? Para comenzar, es necesario *a priori* categorizar los datos como información pública o privada. La información privada, *a prima facie* es aquella sobre personas, la cual debe mantenerse confidencial, y las organizaciones deben utilizar controles de seguridad adecuados para cumplir con estas reglas (Kim & Solomon, 2021, 42), y de hecho, desde esta concepción son equivalentes a datos personales. Por esta razón, se dejará por un lado la noción de datos privados

como datos personales (la cual ya fue tratada), para poder abordar los datos públicos en general y su inverso, datos privados, pero en el sentido de datos propietarios.

Verbigracia, los datos en movimiento deben ser protegidos mediante cifrado durante su transferencia a través de redes para prevenir accesos no autorizados. Los datos en tiempo real, que son críticos para análisis inmediatos, requieren controles estrictos de acceso y autenticación para evitar interrupciones o usos indebidos. Los datos en tránsito, aunque solo pasan por un territorio sin ser procesados, deben estar respaldados por evaluaciones de impacto de transferencias para asegurar un cumplimiento legal. Los datos en reposo, que son almacenados en sistemas estables, requieren medidas de cifrado y políticas de retención claras para garantizar su integridad y confidencialidad. Finalmente, los datos en uso, que son temporales y descifrados en memoria, deben estar protegidos con controles de acceso y técnicas avanzadas como cifrado homomórfico en entornos sensibles como la computación en la nube.

2.1.1.7 Datos públicos, privados e históricos

Para comenzar, cuando se mencionan **datos públicos**, existen 3 interpretaciones, la primera, es hacia aquellos **datos de fuentes de acceso público**; segundo, **datos de dominio público**; y tercero, **datos abiertos**.

Respecto a la primera idea, los datos de fuentes públicas, aunque estén disponibles para todos, pueden necesitar consentimiento y una base legal para su procesamiento, especialmente si contienen información personal. En el caso de datos no personales, su disponibilidad pública no garantiza necesariamente el derecho a usarlos para ciertos propósitos, como en el caso de invenciones basadas en patentes en registros públicos. Aquí aplica la regla de si está al acceso público, no significa que puedas utilizarlo.

Sobre la segunda concepción relativa a datos de dominio público son datos que no están protegidos por derechos de propiedad intelectual u otros derechos con efectos similares, y por lo tanto, se encuentran en el dominio público

(entendido de manera más amplia que simplemente estar libre de protección por derechos de autor), por lo tanto, son libres de acceder y reutilizar (Burdon, 2020, 28). En este sentido, los datos en el dominio público pueden ser considerados como un bien público ya que carecen de derechos de propiedad intelectual, lo que permite a cualquier persona compartirlos o distribuirlos libremente (Ploeger & Van Loenen, 2018, 277). Por ejemplo, una patente publicada en un registro público pasa al dominio público, y cualquiera puede usarla después de que expire el período de vigencia de la patente (normalmente 20 años).

Respecto a la tercera interpretación, está el termino datos abiertos. Estos son aquellos que están disponibles y, según su licencia, cualquier persona puede acceder, utilizar, modificar y compartir, sujeto únicamente, como máximo, a los requisitos de proporcionar atribución y compartir de manera similar (Open Data Handbook, s.f.). Asimismo, estos datos abiertos deben cumplir con los siguientes criterios: ser accesible (idealmente a través de Internet) a un costo que no exceda el de la reproducción, sin limitaciones basadas en la identidad o intención del usuario; estar en un formato digital legible por máquina para interoperabilidad con otros datos; y frecuentemente libre de restricciones de uso o redistribución en sus condiciones de licencia (ICO, 2012). Se debe hacer la aclaración que, aunque los datos abiertos podrían ser sinónimos de datos de dominio o fuentes públicas de información (Charles & Emrouznejad, 2019, 3), esto es solo parcialmente exacto porque la naturaleza de los datos abiertos es que están en un formato estructurado, legible por máquina y reutilizable sin restricciones. En un sentido estricto, los datos públicos están en todas partes, y su categoría solo se otorga por ser publicados como fuente pública de información. Es decir, pueden ser datos sin un formato estructurado, y se desconoce su requisito para su reutilización.

Ahora bien, dejando el debate de los datos públicos por un lado, y con un enfoque en los datos privados, más allá de los datos personales, es decir, a datos propietarios, se precisa que, los **datos privados** (propietarios) son un activo que abarca todos los datos propietarios que suelen estar protegidos por Propiedad Intelectual (incluidos los secretos comerciales) u otros derechos de acceso y control (proporcionados por contrato y la ley de ciberseguridad), para los cuales

generalmente existe un interés económico en excluir a otras personas del uso de los activos de propiedad intelectual sin el consentimiento del titular (OCDE, 2019a).

Dicho lo anterior, la distinción entre datos privados (propietario) y públicos enfrenta varios desafíos debido a los avances tecnológicos. En primer lugar, los datos del sector público y privado no siempre pueden distinguirse, por ejemplo, los datos generados, creados, recopilados y procesados por el sector privado financiado por o para el sector público se clasificarían como datos del sector público y privado; en segundo lugar, es necesario tener una mayor claridad entre los datos pertenecientes al sector público y aquellos del sector privado, ya que con frecuencia se confunden erróneamente con los datos del dominio público y los datos privados propietarios, respectivamente; los datos creados y gestionados por el sector público suelen considerarse como propietarios antes de ponerlos a disposición del público, y por el contrario, los datos privados o propietarios pueden pasar al dominio público si se abren; en tercer lugar, la distinción entre datos del sector privado y del sector público no refleja completamente los datos domésticos, que son principalmente datos personales (OCDE, 2019a). En conclusión, es imperativo establecer reglas para diferenciar entre datos en el dominio público y privado, ya que su tratamiento depende de su clasificación y tratamiento.

Prosiguiendo en el análisis. Es necesario mencionar brevemente un tipo de dato, que si bien no se le podría poner atención por su temporalidad, son necesarios para construir el futuro; los **datos históricos**. De manera puntual, los datos históricos representan y encarnan la vida cultural humana (Cameron, 2021, 28). Aunque los datos históricos no suelen considerarse patrimonio, pueden clasificarse como tal si se incluyen en archivos patrimoniales y están regidos por regulaciones y procedimientos institucionales. Sin embargo, en el régimen actual, todos los datos se perciben como encarnación del pasado. Además, son esenciales para el futuro; por ejemplo, las predicciones financieras se basan en datos históricos, donde los datos de riqueza se procesan para prever la vida futura en todas sus posibles formas (Cameron, 2021, 80-88).

Otros subconceptos para los datos históricos son **datos fijos**, **permanentes**, **de referencia**, **archivísticos** o de **contenido fijo**; son datos que normalmente no están sujetos a cambios. Por ejemplo, los datos fijos incluyen resultados de investigaciones concluidas, registros médicos y datos históricos (TechTarget, s.f.). Según el Considerando 156 del RGPD, el procesamiento de datos personales con fines de archivo en interés público, investigación científica o histórica, o fines estadísticos debe estar sujeto a garantías adecuadas para los derechos y libertades de las personas interesadas. Los datos históricos pueden ser datos personales, dependiendo de la información específica que contengan y si esa información se relaciona con individuos identificables. Si los datos se consideran datos personales o no, se determina principalmente por si pueden identificar a una persona directa o indirectamente.

Los datos públicos incluyen datos de fuentes públicas, datos en el dominio público y datos abiertos, cada uno con diferentes niveles de acceso y restricciones. Por ejemplo, los datos abiertos deben estar estructurados y ser reutilizables sin limitaciones, mientras que los datos de fuentes públicas, aunque accesibles, pueden requerir una base legal para su uso. Los datos privados, especialmente los propietarios, están protegidos por derechos de propiedad intelectual o acuerdos contractuales y requieren controles estrictos para prevenir usos no autorizados. Por último, los datos históricos, esenciales para la investigación y la preservación cultural, deben manejarse bajo garantías adecuadas, asegurando que su procesamiento no comprometa los derechos y libertades individuales, especialmente cuando contienen información personal.

2.1.1.8 Datos íntimos, confidenciales, restringidos, prohibidos y clasificados

Llegados a este punto, es imperante abordar algunos datos que están relacionados con su nivel de apertura y restricción. Estos datos son: **datos íntimos, datos confidenciales, datos restringidos, datos prohibidos** y **datos clasificados**.

Inicialmente, los **datos íntimos** son parte de la vida que se quiere mantener reservada y tener pleno control sobre ellos. Con este fin, se protege la no divulgación de datos de la vida privada de una persona, es decir, que otros no conozcan aspectos de su vida sin su consentimiento y se integra con los aspectos más personales de la vida y el entorno familiar, cuyo conocimiento está reservado para los miembros de la unidad familiar (Ferrer, 2019, 459-460). Así pues, al separar la protección de datos, la privacidad y la intimidad, se puede tener más luz sobre los bienes jurídicos que se protegen. Cabe destacar que no existe derechos absolutos, y por ello diversas regulaciones contienen excepciones justificables a la intervención en comunicaciones privadas bajo orden escrita de una autoridad competente o el interés público, colocando otros derechos encima a la privacidad (en este caso de comunicaciones). Sin embargo, debe quedar claro que la intimidad es un derecho inmutable e inviolable.

Por otra parte, los **datos confidenciales** se refieren a aquellos datos protegidos contra la divulgación de información asegurándose de que los datos estén limitados a aquellos autorizados de tal manera que su acceso esté disponible solo para aquellos que poseen alguna información crítica, por ejemplo, una clave para descifrar los datos cifrados (NIST, s.f.). En el contexto de datos personales, los datos confidenciales son cualquier información que pueda identificar directa o indirectamente a un individuo y potencialmente estar bajo los términos de legislación especial de protección de datos. Por ejemplo, información sobre aspectos fiscales, bancarios, fiduciarios, o cualquier otro, según disposición legal expresa, cuya propiedad corresponde a individuos.

Por su parte, los **datos restringidos** son aquellos relacionados con (i) el diseño, fabricación o utilización de armas nucleares; (ii) la producción de material nuclear; o (iii) el uso de material nuclear en la producción de energía, pero no incluirán datos desclasificados o retirados de las restricciones (NIST, s.f.). A tal efecto, el Acta de Energía Atómica de los Estados Unidos considera los secretos nucleares como datos restringidos, los cuales operan de manera diferente a todas las demás formas de clasificación de seguridad nacional y que aún existen en la actualidad (Wellerstein, 2021, 154). Sin embargo, este término, que se aplicaba a

toda la información nuclear, no distingue la naturaleza de la información, ya que no importaba quién generara los datos ni con qué propósito (Wellerstein, 2021, 300), además por razones lógicas, estos datos no eran publicables y no tenían clasificaciones, es decir, a diferencia de los grados militares de clasificación, no sería posible saber si esta información era secreta, ultra secreta o confidencial (Wellerstein, 2021, 194).

La regulación del Acta de Energía Atómica fue pionera en la implementación del lenguaje legal de los delitos informáticos hace más de 60 años, contemplando datos restringidos y personas no autorizadas; un hecho importante fue cuando, en 2015, un ex científico del Laboratorio Nacional de Los Álamos fue condenado a 60 meses de prisión por violar la cláusula de datos restringidos del Acta de Energía Atómica de los EE. UU. (Schreider, 2020, 38). En *stricto sensu*, los datos restringidos son aquellos catalogados por la legislación, como por ejemplo el Acta de Energía Atómica, pero en un sentido más amplio, los datos restringidos pueden interpretarse como datos propietarios y exclusivos para personal autorizado, cuya disposición puede requerir permisos legales (Sharma & Pandey, 2020, 62). Actualmente, los datos restringidos podrían compartirse bilateralmente mediante un acuerdo, como por ejemplo a través de un memorando de entendimiento (World Bank Group, 2021).

Se debe hacer la aclaración que los datos restringidos y los datos personales no son necesariamente lo mismo, pero pueden superponerse dependiendo del contexto y las regulaciones vigentes. Los datos restringidos generalmente se refieren a datos sujetos a ciertas restricciones o limitaciones en su uso, acceso o divulgación. Estas restricciones pueden surgir de diversas fuentes, como requisitos legales, acuerdos contractuales y políticas organizativas. En algunos casos, los datos personales también pueden clasificarse como datos restringidos, especialmente cuando están sujetos a restricciones legales o contractuales específicas.

Otro tipo de datos vinculados a un nivel de restricción son los **datos prohibidos**. Estos son información de alto secreto, información de seguridad de comunicaciones excluyendo elementos criptográficos controlados cuando no estén

cifrados y se utilicen con claves no clasificadas, datos restringidos, información de programas de acceso especial o información sensible compartimentada (NIST, s.f.). Los datos prohibidos varían y dependen del tiempo y el espacio. Por ejemplo, algunos datos como las opiniones políticas y religiosas y el comportamiento sexual estaban prohibidos (Committee on Government Operations, 1975, 58). Actualmente, se puede inferir que existen datos prohibidos, en la Ley de Inteligencia Artificial de la UE (2024), que en su el artículo 5 prohíbe el uso de sistemas de identificación biométrica remota en tiempo real en espacios públicamente accesibles, con algunas excepciones; si bien prohíbe los sistemas de identificación biométrica, indirectamente se prohíbe la obtención de datos biométricos en tiempo real, sin menoscabo de sus excepciones.

En algunas jurisdicciones, se delimitarán las reglas de acciones y tipos de datos prohibidos, por ejemplo, la prohibición de procesar datos de menores, datos íntimos o intervenciones en comunicaciones sin una orden judicial.

Por otra parte, se abordan los **datos clasificados**. Desde una concepción estricta, los datos clasificados o información clasificada de seguridad nacional significan información que ha sido determinada según orden o cualquier orden anterior que requiera protección contra la divulgación no autorizada y está marcada para indicar su estado clasificado cuando está en forma documental (NIST, s.f.).

Desde una concepción amplia, los datos clasificados son información sensible designada con niveles de seguridad específicos según su importancia y riesgo. Se manejan conforme a normativas de seguridad y solo están disponibles para personas autorizadas.

Al respecto, en este sentido existen varios niveles de clasificación gubernamental o militar: primero, los **datos de máximo secreto**, el nivel más alto de clasificación, donde la divulgación no autorizada tendrá efectos drásticos y causará un grave daño a la seguridad nacional; segundo, los **datos secretos**, cuya naturaleza es restringida y la divulgación no autorizada tiene efectos significativos, causando un daño crítico a la seguridad nacional; tercero, los **datos confidenciales** que se utilizan para información sensible, confidencial o altamente

valiosa, cuya divulgación no autorizada tendrá efectos significativos y causará daños a la seguridad nacional; cuarto, **los datos sensibles pero no clasificados**, que son para uso interno o solo de oficina, y tienen como objetivo proteger la privacidad de las personas; y quinto, los **datos no clasificados**, que no son confidenciales ni clasificados, cuya divulgación no compromete la confidencialidad ni causa ningún daño perceptible; además, todos los datos clasificados están exentos de la Ley de Libertad de Información, así como de muchas otras leyes y regulaciones (Chapple et al., 2018, 21-22).

Por ejemplo, los datos íntimos se relacionan con aspectos privados de la vida personal y requieren un control estricto, asegurando que su divulgación sólo ocurra con la divulgación explícita de la persona. Los datos confidenciales, protegidos por su relevancia para la seguridad o privacidad, deben mantenerse restringidos a personal autorizado y protegidos mediante cifrado y medidas de acceso controlado. Los datos restringidos, como información nuclear o de seguridad nacional, requieren regulaciones específicas, acuerdos bilaterales, y evaluaciones rigurosas para evitar accesos no autorizados. Los datos prohibidos, que incluyen información cuya recolección o uso está vetado por ley, como datos biométricos en tiempo real en ciertas jurisdicciones, deben ser manejados con un estricto apego a las restricciones legales. Finalmente, los datos clasificados, jerarquizados en niveles de sensibilidad (máximo secreto, secreto, confidencial), deben estar sujetos a protocolos rigurosos de acceso, manejo y almacenamiento, garantizando que solo personal autorizado con la debida acreditación pueda acceder a ellos, en línea con normativas de seguridad nacional.

2.1.1.9 Datos cifrados y no cifrados

Finalmente, una última clasificación conforme a un nivel técnico de seguridad son los **datos cifrados** y **datos no cifrados**. Respecto al primer término, datos no cifrados, también se les puede conocer mediante el anglicismo de datos encriptados, la cual es un método o proceso para proteger la información de ataques no deseados al convertirla en una forma que resulta irreconocible para los

atacantes; la encriptación de datos consiste en convertir datos específicos, como texto, imágenes, audio y videos, en una forma ilegible o invisible durante la transmisión (Shaheen, 2021, 6).

La encriptación es una necesidad absoluta en el procesamiento de datos de las regulaciones de protección de datos, siendo una herramienta fundamental para proteger datos personales o, en el peor de los casos, ocultar información de actividades criminales. Por ejemplo, la controversia en la que el gobierno de EE. UU. solicitó a Apple crear una llave maestra o puerta trasera para la encriptación colocada en los productos *iPhone* para que las fuerzas del orden pudieran acceder a los archivos de un terrorista fallecido (Yadron et. al, 2016). Sin embargo, sobre este tema de famosas llaves maestras en manos del gobierno, la UE ha declarado que este tipo de llaves y soluciones representan más una amenaza para la sociedad que acciones beneficiosas (Consejo de la Unión Europea, 2020).

Otra forma de definir los datos encriptados es como **datos en negro (*black data*)**, siendo aquellos que están protegidos por encriptación para ser transportados o almacenados sin temor a comprometer su seguridad, también conocidos como datos encriptados (NIST, s.f.). O también como **datos oscurecidos (*obscured data*)**, que son datos que han sido distorsionados mediante criptografía u otros medios para ocultar información y también se les conoce como datos enmascarados u ofuscados (NIST, s.f.).

Y aunque los datos cifrados sean una medida robusta en seguridad de la información. En teoría, los algoritmos de computación cuántica podrían, de hecho, causar estragos en el mundo de la seguridad de datos (cifrado) tal como lo conocemos hoy en día (Krishnakumar, 2020, 257). Pero indistintamente de la concepción de que nada es seguro y con una visión propositiva, la realidad es que los datos encriptados son útiles para la privacidad y la seguridad de la información. Al respecto, existen avances tecnológicos que han mejorado algunas técnicas que permiten el uso de datos encriptados. Por ejemplo, la encriptación completamente homomórfica es un método de criptografía que permite sumar y multiplicar datos encriptados y su integración en algoritmos de aprendizaje automático todavía está en sus primeras etapas, pero sugiere que el aprendizaje sobre datos encriptados

podría ser una estrategia razonable cuando la sensibilidad de los datos es alta (Hamon et al., 2020, 21).

Por lo tanto, si se utiliza la encriptación como medida técnica para asegurar los datos, es vital garantizar la seguridad de la clave; en consecuencia, es obligatorio contar con un sólido sistema de gestión de claves para mantener el nivel de protección que ofrece el mecanismo de encriptación (ICO, 2013). Empero, el objetivo a largo plazo no es solo proteger los métodos criptográficos existentes contra la computación cuántica, sino eventualmente crear nuevos tipos de sistemas criptográficos adaptados a la era cuántica (Swan et al., 2020, 177).

A *contrario sensu*, de los datos cifrados, se encuentran los **datos no cifrados**. Estos son datos originales sin encriptar o texto plano (Munir & Mohammed, 2019, 703). Otro concepto inusual contemplado por NIST (s.f.) es el de datos RED (datos que no están protegidos por encriptación, también conocidos como datos sin encriptar). Desde el aspecto jurídico, se debe subrayar que, en algunas jurisdicciones, existe la obligación legal de proporcionar las claves de acceso. Además de atacar la encriptación, las autoridades policiales y judiciales también pueden intentar eludir la encriptación solicitando u ordenando la entrega de los datos sin encriptar o la clave de acceso (Europol, 2020).

Finalmente, en cuanto a los datos encriptados y sin encriptar, es importante tener en cuenta que en casos de violación de seguridad (se considera que ha ocurrido una violación cuando una organización descubre que un tercero ha obtenido acceso no autorizado a la información) existe la responsabilidad de activar la notificación de violación cuando una organización tiene motivos razonables para creer que ocurrió un acceso no autorizado, incluso si no están completamente seguros, lo que ha llevado a que muchas leyes estatales especifiquen que si los datos en cuestión estaban encriptados o suficientemente redactados, entonces el acceso no autorizado no constituye una violación (Chapple & Shelley, 2021, 213).

Sin embargo, cuando se utiliza de manera desleal, este precepto de únicamente notificar cuando la información ha sido desencriptada podría promover

a ocultar incidentes de seguridad, que si bien no fueron brechas de seguridad si fueron incidentes y más que un sencillo evento o alerta.

En este aspecto, es cierto que la mayoría de las leyes han tratado el cifrado como invulnerable a la divulgación y han eximido todas las divulgaciones de información encriptada de los requisitos de notificación. Pero, la encriptación es inútil si el atacante obtiene las claves de encriptación. Por ejemplo, en 2016, el proyecto de ley SB 2005 de Tennessee actualizó las reglas de notificación de violaciones para incluir incluso datos encriptados en la definición de una violación de información personal, y para evitar los requisitos de notificación de violaciones, las organizaciones deben demostrar que los datos fueron encriptados según un estándar federal específico y que las claves de encriptación no han sido comprometidas (Chapple & Shelley, 2021, 215). Por lo tanto, será necesario encontrar soluciones adecuadas para incidentes de seguridad y datos encriptados comprometidos.

Por ejemplo, los datos cifrados requieren algoritmos robustos, gestión segura de claves y preparación frente a riesgos futuros, como la computación cuántica. Tecnologías como el cifrado homomórfico permiten operar sobre datos cifrados en contextos sensibles. Los datos en texto plano o no cifrados, al presentar un mayor riesgo de exposición directa si no se implementan medidas de protección adecuadas, exigen controles estrictos de acceso, monitoreo continuo y evaluaciones de seguridad. Tanto datos cifrados como no cifrados están sujetos a regulaciones que pueden exigir notificaciones de violaciones. Es esencial complementar el cifrado con estrategias integrales de seguridad, reconociendo que no es una solución absoluta.

2.1.1.10 Datos del producto, de servicios relacionados, fácilmente disponibles y exportables

El Reglamento de Datos (2023), establece normas armonizadas para un acceso equitativo a los datos y su utilización, modificando reglamentos previos y define

nuevos tipos de datos, siendo: Datos del producto, de servicios relacionados, fácilmente disponibles y exportables

En primer lugar, se definen los **datos del producto** como aquellos generados por el uso de un producto conectado, diseñado para ser extraídos por el usuario, tenedor de los datos o un tercero. Reconoce la importancia de estos datos en la economía digital y busca eliminar obstáculos al mercado interior de datos. Se especifica que antes de la compra o alquiler de un producto conectado, el vendedor debe informar claramente al usuario sobre los datos que puede generar el producto, incluyendo su tipo, formato y volumen. Establece que el fabricante, como tenedor de datos, puede utilizarlos basándose en un contrato con el usuario. Además, establece que los datos del producto solo deben ponerse a disposición de terceros con el consentimiento del usuario y complementa el derecho de los interesados a recibir y transmitir datos personales según el RGPD.

En segundo lugar, los **datos de servicios relacionados** son aquellos que representan la digitalización de acciones del usuario o eventos asociados con un producto conectado, siendo registrados intencionadamente por el usuario o generados como resultado de su actividad durante la prestación de un servicio por parte del proveedor. Estos datos deben ser accesibles para el usuario de manera fácil, segura y gratuita, en un formato completo, estructurado y de uso común, así como, cuando sea posible, directamente accesibles por el usuario.

En tercer lugar, los **datos fácilmente disponibles** se refieren a la información del producto y del servicio relacionado que un tenedor de datos puede obtener legalmente sin un esfuerzo desproporcionado, como parte de una operación simple. Esto puede incluir datos obtenidos a través del diseño del producto, contratos entre el tenedor de datos y el usuario para servicios relacionados, y medios técnicos de acceso. Estos datos no abarcan la información generada por el uso del producto si el diseño no prevé su almacenamiento o transmisión fuera del producto en sí. Además, se establecen disposiciones para compartir estos datos con terceros bajo ciertas condiciones y con garantías de calidad y seguridad.

En cuarto lugar, los **datos exportables** se refieren a los datos de entrada y salida, junto con los metadatos, generados directa o indirectamente por el uso del servicio de tratamiento de datos por parte del cliente, excluyendo activos o datos protegidos por derechos de propiedad intelectual o secretos comerciales de terceros. Para facilitar la transición del cliente a otro proveedor de servicios de tratamiento de datos, el proveedor original debe informar al cliente sobre qué datos puede exportar, garantizando que no pierda ninguno de sus datos durante el proceso de cambio. Los datos exportables deben incluir, como mínimo, los datos mencionados anteriormente, excluyendo aquellos protegidos por derechos de propiedad intelectual o secretos comerciales, así como datos relacionados con la seguridad cuya exportación pueda exponer a vulnerabilidades de ciberseguridad, sin que estas exclusiones obstaculicen el proceso de cambio.

En resumen, el Reglamento de Datos (2023) introduce normas armonizadas para el acceso equitativo y la utilización de diferentes tipos de datos, que incluyen Datos del Producto, Datos de Servicios Relacionados, Datos Fácilmente Disponibles y Datos Exportables. Los Datos del Producto son generados por el uso de productos conectados y pueden ser extraídos por el usuario o un tercero, con el fabricante como tenedor de datos. Se requiere que el vendedor informe al usuario sobre los datos generados por el producto antes de la compra o alquiler. Los Datos de Servicios Relacionados representan la digitalización de acciones del usuario o eventos asociados con un producto conectado, accesibles fácil, segura y gratuitamente para el usuario. Los Datos Fácilmente Disponibles son aquellos que pueden obtenerse legalmente sin esfuerzo desproporcionado y se comparten con terceros bajo ciertas condiciones. Finalmente, los Datos Exportables son los datos de entrada y salida generados por el uso del servicio de tratamiento de datos por parte del cliente, excluyendo activos o datos protegidos por derechos de propiedad intelectual o secretos comerciales de terceros, facilitando la transición del cliente a otro proveedor de servicios de tratamiento de datos. Estos tipos de datos se complementan para garantizar un acceso adecuado y una transición fluida entre proveedores de servicios de tratamiento de datos.

Por ejemplo, los datos del producto, generados por productos conectados, deben gestionarse respetando el consentimiento del usuario, garantizando que estos puedan acceder, extraer o compartir dichos datos sin obstáculos indebidos. Los datos de servicios relacionados, que digitalizan interacciones del usuario con un servicio, deben ser accesibles de manera gratuita, estructurada y segura, alineándose con los principios de usabilidad y control del usuario. Los datos fácilmente disponibles deben ser obtenidos y compartidos sin esfuerzos desproporcionados, asegurando su calidad y cumplimiento de condiciones contractuales y técnicas. Finalmente, los datos exportables, que son clave para la interoperabilidad y la transición entre proveedores, deben garantizar que el cliente pueda transferir su información sin pérdida ni exposición a riesgos de ciberseguridad, respetando exclusiones como secretos comerciales y derechos de propiedad intelectual.

2.1.2 Nivel Buenas Prácticas

En el complejo mundo de gestión de datos y cumplimiento normativo, es necesario explorar el segundo nivel de la teoría propuesta: el nivel de Buenas Prácticas. En este apartado, se abordará más allá del ámbito de las reglas y principios legales vinculantes, adentrándose en un dominio donde los datos asumen un significado diferente. En este nivel, la ley no obliga directamente al cumplimiento, sino que se guía por el conocimiento y la previsión.

El nivel de Buenas Prácticas ofrece una perspectiva matizada, donde los datos adquieren un papel más allá de la mera obligación. En cambio, asumen el manto de la responsabilidad y la administración ética. Aunque puede que no tenga el peso de la ejecución legal, adherirse a estos principios es una evidencia del compromiso de una organización con la gestión ética y responsable de los datos.

Extrayendo su esencia de fuentes a menudo denominadas como derecho indicativo o *soft law*, ya sea mediante los códigos de conducta de la industria, los principios de privacidad y los marcos de protección de datos, los estándares de Buenas Prácticas trazan un camino de autocontrol e integridad. Las

organizaciones que abrazan estos estándares se distinguen como defensores de las mejores prácticas en el manejo de datos.

Sin más preámbulo, esta exploración profundizará en el mundo de las Buenas Prácticas, desentrañando las sutilezas y mostrando cómo la adhesión a estos principios puede mejorar la confianza, la reputación y mitigar los riesgos que a menudo acompañan a un manejo inadecuado de ciertos datos.

Esta categoría de nivel buenas prácticas, incluye datos volátiles y no volátiles; Datos dinámicos y estáticos; Datos de entrada y de salida (input - output); Datos generados por máquinas; Datos de internos y externos; Datos estructurados; semiestructurados; y no estructurados.

2.1.2.1 Datos volátiles y no volátiles

El primer dato que tratar en este apartado son los **datos volátiles**. Estos datos son aquellos que se encuentran en un sistema en funcionamiento y se pierden cuando se apaga el ordenador (NIST, s.f.). Las regulaciones de protección de datos principalmente protegen los datos en almacenamiento, procesamiento y tránsito; los datos no se almacenan permanentemente, pero pueden ser procesados en este caso. Por lo tanto, las regulaciones de datos se aplican a ellos. Además, es crucial evitar perder datos volátiles, ya que podrían servir como evidencia confiable para procesos jurisdiccionales, por ejemplo, los examinadores deben saber cuándo se deben recopilar estos datos y cómo hacerlo (Lyn, 2020, 430).

En ciertas circunstancias, es vital considerar los datos volátiles. Por ejemplo, acceder a un archivo requiere una copia temporal en la memoria en casos que involucran propiedad intelectual. Del mismo modo, en el caso de ciberdelitos, alguien puede tener permiso para copiar un archivo en la memoria, pero no en el disco duro, lo que dificulta determinar si ha ocurrido una acción no autorizada y si constituye un delito penal según las leyes de un país específico (Lambert, 2022, 313-317).

Por el contrario, los **datos no volátiles** son aquellos que persisten incluso después de que se apaga un ordenador (NIST, s.f.). Estos tipos de datos suelen ser la fuente más grande y valiosa de evidencia digital potencial que puede ser analizada durante una investigación forense.

En cuanto a la informática forense, los registros recuperados de un ordenador pueden considerarse rumores, dependiendo de cómo fueron creados inicialmente, pero en contraste, muchos tribunales han sostenido que los registros de datos generados por máquina, como los registros y la salida de programas informáticos, no son rumores porque estos registros se crean sin intervención humana, y por lo tanto, esa información generada por computadora no es una declaración de una persona y no puede ser considerada como rumores, por lo tanto, los registros son admisibles (Lyn, 2020, 441)

Por ejemplo, los datos volátiles, presentes en la memoria activa y perdidos al apagar un sistema, requieren captura rápida y precisa, especialmente en investigaciones forenses, ya que pueden contener evidencia como actividad reciente o datos en tránsito, y su manejo debe realizarse con herramientas especializadas para preservar su integridad y admisibilidad legal. Por otra parte, los datos no volátiles, almacenados de forma persistente en discos o dispositivos, son la principal fuente de evidencia digital y deben analizarse detalladamente, garantizando su autenticidad mediante cadenas de custodia adecuadas.

2.1.2.2 Datos dinámicos y estáticos

Por otro lado, los **datos dinámicos** son aquellos que pueden cambiar durante su vida útil (Butterfield et al., 2016). Por ejemplo, la información sobre el clima; puede llover hoy y no mañana, y la probabilidad de lluvia cambia, lo que afecta cómo alguien planea actividades al aire libre (Mueller & Massaron, 2019, 121). Y por ello, en los datos dinámicos, la vinculación y la autenticación son un gran desafío; podría ser necesario agregar una marca de tiempo en cada registro que indique cuándo fue creado o actualizado por última vez (Christen et al., 2020, 70).

Es imperativo tener en cuenta que, con la aparición de nuevas tecnologías de la información y la comunicación, se debe considerar que los datos son una entidad de continuidad, ya que las fuentes de datos están constantemente actualizadas y cambiando (Hurwitz et al., 2020, 29-46). El caso extremo de datos dinámicos corresponde a aquellos que cambian en tiempo real, lo que hace imposible cumplir con el principio de precisión que rige las regulaciones sobre protección de datos personales.

En contraste, los **datos estáticos** son datos que nunca cambian (Mueller & Massaron, 2019, 121). Por ejemplo, los datos biológicos incluyen datos estáticos, como estructuras de proteínas, patrones de expresión génica medidos con *microarrays* y redes regulatorias, entre otros (Lesk, 2014, 288-291). Otro ejemplo de datos estáticos es un antiguo mapa de Francia en 1950, mientras que otros podrían preferir datos en tiempo real, como el flujo de tráfico actual en el centro de la ciudad (Van Loenen, 2018, 37).

Aunque, los datos históricos no son el único tipo de datos estáticos; por ejemplo, la tecnología *blockchain* tiene datos estáticos almacenados en el conjunto de datos de la red, e incluso el código fuente ejecutable puede persistir en la cadena de bloques sin eliminación o modificación (Bambara & Allen, 2018, 5). Estos datos también son conocidos como **datos persistentes** porque las estructuras de datos que no cambian se llaman inmutables o persistentes (Haverbeke, 2018, 121). En informática forense, los datos persistentes se almacenan en un medio de almacenamiento que los preserva cuando un dispositivo electrónico está apagado (Lyn, 2020, 429). En el caso de la IA, las redes neuronales a veces pueden ser utilizadas para la minería de datos en una gran cantidad de conjuntos de datos estáticos, centrándose en la clasificación y el reconocimiento de patrones (Gilchrist, 2016, 58).

Por ejemplo, los datos dinámicos, que cambian continuamente durante su ciclo de vida, como la información climática o los datos en tiempo real, requieren estrategias para garantizar su autenticidad y precisión, como el uso de marcas de tiempo y registros de actualización. Estos datos presentan desafíos únicos para cumplir con regulaciones como el principio de precisión en la protección de datos

personales, especialmente cuando los cambios son constantes. Por otro lado, los datos estáticos, que permanecen inalterados a lo largo del tiempo, como estructuras biológicas, mapas históricos o registros almacenados en *blockchain*, requieren medidas de preservación para garantizar su integridad y disponibilidad a largo plazo.

2.1.2.3 Datos de entrada y de salida (input - output)

Si bien ya se mencionó el uso de sistemas de IA y datos, existe una clasificación más específica de datos para este tipo de sistemas, siendo los **datos de entrada**. En este tenor, de acuerdo con el Artículo 3 (32) de la Ley de Inteligencia Artificial (2024), los datos de entrada se refieren a los datos proporcionados o directamente adquiridos por un sistema de IA en función de los cuales el sistema produce una salida. Y, por lo tanto, es vital verificar la calidad de los datos de entrada de los componentes de IA y que la precisión de los datos sea correcta, ya que pueden ocurrir errores debido a elementos externos a la IA, como lectores biométricos, que introducen errores en la entrada de datos, además, al ejercer el derecho de acceso, es necesario permitir que los titulares conozcan los datos de entrada y salida esperados (REDIPD, 2019).

En cambio, los **datos de salida** son los datos generados por el sistema, incluyendo los datos producidos a nivel de software, como el resultado de un cálculo, o a nivel físico, como un documento impreso (Techterms, s.f.). En algunos casos, los datos de salida también pueden requerir una explicación, especialmente cuando el destinatario de la decisión ha sido colocado en una categoría que puede no ser clara para ellos (Leslie & Briggs, 2021, 16). En resumen, los datos de salida en sí mismos no suelen clasificarse como datos personales, pero podrían contener información sensible o combinarse con otros datos que puedan revelar información personal o identificar indirectamente a individuos. Es fundamental manejar los datos de salida de manera responsable y de acuerdo con los principios de protección de datos y privacidad, especialmente al tratar información potencialmente sensible.

Por ejemplo, los datos de entrada, que alimentan al sistema de IA, deben ser cuidadosamente validados para evitar errores que puedan afectar los resultados. Además, es esencial garantizar la transparencia al permitir a los interesados conocer los datos de entrada utilizados en el sistema, cumpliendo con derechos como el de acceso. Por otro lado, los datos de salida, generados por el sistema, deben gestionarse con responsabilidad, ya que, aunque en principio no suelen ser personales, podrían contener información sensible o combinarse con otros datos que revelen información personal. Adempás, el manejo de los datos de salida debe incluir explicaciones claras, especialmente en decisiones automatizadas que puedan impactar a los usuarios, asegurando que se respeten los principios de privacidad y protección de datos.

2.1.2.4 Datos generados por máquinas

Será preciso mostrar que, hay más datos generados por máquinas que datos generados por humanos (Raj & Raman, 2017, 216). Por esta razón es importante hacer una clasificación para los **datos generados por máquinas**, siendo datos creados automáticamente sin intervención humana (Hurwitz, et al., 2016, 26), por ejemplo, datos creados por una computadora, un proceso, una aplicación u otros medios automáticos (Cielen et al., 2016, 6). Los datos generados por máquinas sin un autor humano pueden plantear problemas de derechos de autor, ya que no hay una propiedad clara de la obra. Por ejemplo, los informes periódicos producidos mediante la ejecución de un script pueden caer en esta categoría, dado que no hay aportación intelectual por parte del autor, el estado de los derechos de autor de tales obras aún está por determinarse (Reed, 2021, 181).

En síntesis, aunque los datos generados por máquinas en sí mismos no son datos personales, pueden vincularse indirectamente a datos personales en ciertas situaciones. Por ejemplo, supongamos que los datos generados por máquinas están asociados con una cuenta de usuario específica o un dispositivo. En ese caso, puede relacionarse indirectamente con datos personales porque se puede rastrear hasta un individuo identificable. Las regulaciones de protección de datos y

privacidad pueden aplicarse en tales casos, dependiendo del contexto y la información específica involucrada. Y a tal efecto, las organizaciones que manejan datos generados por máquinas deben ser conscientes del potencial de identificación indirecta y asegurarse de manejar dichos datos de acuerdo con las leyes de protección de datos y las mejores prácticas de privacidad para salvaguardar los derechos de privacidad de las personas.

Los datos generados automáticamente, como registros de actividad o informes realizados por *scripts*, presentan desafíos en cuanto a propiedad intelectual al no tener un autor humano definido. Aunque no suelen ser datos personales, pueden vincularse a individuos identificables a través de cuentas o dispositivos, lo que activa obligaciones legales de protección de datos. Por ello, es importante que las organizaciones adopten prácticas sólidas de anonimización y gobernanza de datos para garantizar el cumplimiento de las leyes de privacidad, minimizando riesgos de identificación y protegiendo los derechos de las personas.

2.1.2.5 Datos de internos y externos

El siguiente tipo de datos, son los **datos internos**. Los datos internos no tienen el mismo contenido que su versión final; esto podría generar problemas específicos de inconsistencias y confundir a las partes interesadas, lo que dificultará determinar si un borrador es información pública o es un nuevo tipo de datos que no cumple con la ley de transparencia y acceso a la información pública, además, los datos internos son necesarios para desarrollar experimentos y pruebas de concepto o pilotos y sus costos monetarios son bajos (Corea, 2019, 11). Las organizaciones deben reconocer y manejar adecuadamente cualquier dato personal que recolectan o generen como parte de sus operaciones internas, para cumplir en su caso, con las regulaciones de protección de datos y las mejores prácticas de privacidad.

Sin embargo, los datos internos por sí solos no son suficientes, por lo que es vital aprovechar el valor de los **datos externos** y combinarlos con los datos internos para comprenderlos mejor y tener éxito. Por ejemplo, los datos internos

pueden ser históricos con una visión retrospectiva, mientras que los datos externos ayudan a mirar hacia el futuro (Treder, 2020, 302-304). Mientras que los datos internos podrían ser comentarios de consumidores, conversaciones por correo electrónico o notas técnicas, los datos externos podrían provenir de redes sociales, información meteorológica, tráfico, economía o de fuentes externas como datos abiertos (Schmarzo, 2016, 19).

Los datos externos comprenden una variedad infinita de información que existe fuera de una entidad, siendo datos listos para usar de naturaleza pública y privada. Sin embargo, se debe tener en cuenta que los datos externos probablemente tendrán algún costo de acceso y que a veces será arriesgado usarlos debido a la incertidumbre en su precisión (Marr, 2017, 91-93).

En resumen, los datos externos corren el riesgo de tener sesgo, principalmente cuando se obtienen y no se garantiza que sus atributos sean neutrales, por lo que es necesario crear una función central responsable de todas las fuentes de datos externos (Treder, 2020, 38), y para gestionar fuentes de datos externos y estándares, incluyendo acuerdos con proveedores de datos externos (Treder, 2020, 133).

Los datos externos pueden ser datos personales, dependiendo de la naturaleza de los datos y si contienen información relacionada con individuos identificados o identificables. Los datos externos se refieren a datos obtenidos o provenientes de entidades externas, fuentes o terceros en lugar de ser generados o recopilados internamente por una organización. Es fundamental para las organizaciones que recolectan, procesan o utilizan datos externos estar conscientes de la posible presencia de datos personales dentro de esos datos y manejarlos de acuerdo con las regulaciones de protección de datos y las mejores prácticas de privacidad.

En el contexto de *Blockchain*, un oráculo es un sistema que puede responder preguntas externas a la red *Blockchain*, actuando como un tercero que proporciona información externa. Sin embargo, lo que se destaca es la existencia de oráculos descentralizados, los cuales garantizan la disponibilidad de datos y la creación de una red de proveedores de datos individuales con un sistema de

agregación de datos en cadena, lo que permite llegar a un consenso a partir de múltiples fuentes y obtener información más confiable (Antonopoulos & Wood, 2019, 253-265). Lo anterior quiere decir que es necesario consultar fuentes de acceso público para poder hacer un consenso entre múltiples fuentes, pero por encima del consenso descentralizado, lo cierto es que, la revolución industrial ha difuminado el lindero de realidades, y actualmente existe una conexión entre la realidad virtual y analógica, que conlleva a que se deban recopilar datos externos para cumplir con tareas específicas o finalidades.

Los datos internos, generados dentro de una organización, son clave para operaciones y análisis, pero requieren una gestión cuidadosa para evitar inconsistencias y garantizar el cumplimiento regulatorio, especialmente si incluyen datos personales. En contraste, los datos externos, obtenidos de fuentes públicas, privadas o terceros, enriquecen las perspectivas para proyecciones futuras, aunque implican riesgos como sesgos o imprecisiones. Para maximizar su valor, es esencial validar su calidad, gestionar acuerdos con proveedores y garantizar el cumplimiento normativo en caso de contener información personal. Las tecnologías como los oráculos descentralizados en *blockchain* pueden integrar datos externos de forma confiable, demostrando la necesidad de combinar estratégicamente ambas fuentes para obtener resultados precisos y efectivos.

2.1.2.6 Datos estructurados, semiestructurados y no estructurados

Finalmente, a continuación se abordarán tres tipos de datos por su estructura: datos estructurados, semiestructurados y no estructurados. Los **datos estructurados** se refieren a información que tiene una longitud y formato definidos, como números, fechas y nombres (Hurwitz, et al., 2020, 19). Se debe entender que la estructuración de los datos se clasifica como el procesamiento de datos personales; de igual manera, el RGPD cataloga el sistema de ficheros como un conjunto estructurado de datos personales.

Por otra parte, los datos **semiestructurados** son datos que no siguen una estructura estricta, pero muestran una apariencia de lógica en su formato,

generalmente representados por sistemas de clave y valor, donde una clave puede ser entendida como una columna y un valor como una fila en esa columna, de modo que algunas estructuras pueden ser percibidas, pero no una estructura consistente (Simbiose Ventures Team, 2020, 74). Hablando en términos generales, los datos contienen características de identificación y definición, pero no se ajustan a una estructura tan rígida. Un ejemplo son los documentos de texto que contienen datos, generalmente no estructurados, pero se pueden agregar etiquetas o metadatos que representan el contenido y facilitan su procesamiento.

Otros datos que se pueden derivar de datos semiestructurados son los **datos semánticos**, que son datos etiquetados con metadatos particulares que se pueden usar para representar un conjunto de entidades y derivar relaciones, entre otros (Han et al., 2012, 9). Legalmente, esto se considera como datos personales porque contiene identificadores indirectos, por lo que debe tratarse como información personal.

Por último, los **datos no estructurados** son aquellos que no tienen un modelo de datos predefinido y no se ajustan bien a una base de datos relacional tradicional, por ejemplo, carecen de una estructura identificable, incluyendo imágenes, objetos de mapa de bits, texto, audio, video y otros tipos de datos (Isson, 2018, 35-36). Al respecto, utilizando el silogismo de que los datos estructurados son datos personales, a contrario sensu, se podría decir que los datos no estructurados no son datos personales, pero esto es incorrecto.

La premisa anterior es incorrecta porque es probable que gran parte de ese contenido no estructurado incluya datos personales, lo que hace que cualquier procesamiento de dichos datos esté sujeto a las leyes de protección de datos, siempre que el procesamiento sea total o parcial por medios automatizados, o en su efecto, en el caso de tratamiento manual, siempre que los datos procesados formen parte o estén destinados a formar parte de un sistema de archivo; aunque por otra parte, ciertamente el procesamiento manual de archivos o conjuntos de archivos, así como sus portadas, que no estén estructurados según criterios específicos, quedan fuera del alcance material de la protección de datos (Weitzenboeck et al., 2022, 184-206).

Los datos estructurados, organizados en formatos predefinidos como tablas, permiten un procesamiento eficiente y, al contener identificadores, deben cumplir normativas como el RGPD. Los datos semiestructurados, con etiquetas o metadatos, facilitan el análisis pero suelen incluir identificadores indirectos que los clasifican como datos personales bajo leyes de protección de datos. Por su parte, los datos no estructurados, como imágenes o texto libre, carecen de un modelo fijo, pero a menudo contienen información personal regulada al ser procesados. Una gestión efectiva requiere tecnologías de clasificación y análisis que garanticen el cumplimiento normativo, protejan la privacidad y maximicen el valor de cada tipo de dato.

2.1.3 Nivel Consideraciones Futuras

A medida que se avanza en la dinámica ruta de gestión de datos y el cumplimiento normativo, se ha llegado a la última categoría: el nivel de Consideraciones Futuras. En esta sección, se trascienden los límites de las regulaciones vigentes y se abordarán datos desde un enfoque proactivo hacia los territorios inexplorados de la gobernanza de datos del futuro.

Este nivel, aunque aún no está vinculado a las tendencias de regulaciones formales, el surgimiento de nuevos tipos de datos o antiguos datos con renovados debates, demandan la atención y previsión jurídica. Es decir, la sociedad se encuentra al borde de lo que pronto podría ser crucial desde una perspectiva legal y ética. En este sentido, el enfoque proactivo es abordar estos debates y datos emergentes de manera preventiva, preparando a los especialistas para los futuros marcos legales y paradigmas éticos.

En síntesis, en este último nivel de la Teoría de los Tres Niveles de Datos, se pretende que las organizaciones puedan posicionarse como pioneras en anticipar y abrazar el cambiante panorama de datos. Al hacerlo, se mantendrán a la vanguardia y se posicionarán para cumplir con las futuras leyes y regulaciones, teniendo en cuenta los principios de la naturaleza y atributos de los datos ya conocidos y la aparición de novísimos datos.

Esta categoría de nivel consideraciones futuras, incluye Datos mentales; Datos sintéticos; Datos degradados/desvanecidos; y perecederos; Datos fragmentados; y aislados.

2.1.3.1 Datos mentales

Para comenzar, dada la aparición de neurotecnologías, se deriva la necesidad de regular los **datos mentales** o *neurorights*. Por ejemplo, desde un enfoque de *soft law* o derecho indicativo, el Gobierno de España (2021) ha emitido la Carta de Derechos Digitales, reconociendo los derechos digitales en el uso de las neurotecnologías.

Por otro lado, desde la esfera del derecho vinculante o *hard law*, la Constitución Política de la República de Chile (1980), en su artículo 19 de su Constitución Política, reconoce la protección de la información cerebral. Textualmente la Constitución establece lo siguiente:

> CAPÍTULO III DE LOS DERECHOS Y DEBERES CONSTITUCIONALES.
>
> Artículo 19. La Constitución asegura a todas las personas:
>
> 1°. El derecho a la vida y a la integridad física y psíquica de la persona. La ley protege la vida del que está por nacer
>
> (...)
>
> (...)
>
> El desarrollo científico y tecnológico estará al servicio de las personas y se llevará a cabo con respeto a la vida y a la integridad física y psíquica. La ley regulará los requisitos, condiciones y restricciones para su utilización en las personas, debiendo resguardar especialmente la actividad cerebral, así como la información proveniente de ella.
>
> 2°. al 26°. (...)

Teniendo en cuenta todo lo dicho anteriormente, se puede precisar que los *neurorights* se basan en la protección de las ideas y pensamientos de las personas, y que por encima de vincularlos a las regulaciones relativas a protección de datos personales, se relacionan más hacia un derecho a la privacidad. Es decir, la noción de control mental subyace a los principios entre el derecho al silencio y

los valores de libertad y privacidad. Por ejemplo, la imagen cerebral enfatiza la importancia del control mental desde un punto de vista legal y moral; como resultado, el concepto de *neurorights*, o el derecho al control mental, prohíbe al Estado extraer los pensamientos de un sospechoso sin su consentimiento o utilizar su memoria o reconocimiento (Fox, 2011, 335).

Dada la importancia de salvaguardar los datos mentales relacionados con los neuroderechos, surgen interrogantes sobre los bienes jurídicos, derechos y obligaciones asociados a ello. Algunas de estas cuestiones son: ¿Puede obligarse a una persona a proporcionar su clave de cifrado o contraseña? ¿Puede el gobierno obligar a un propietario de datos a proporcionar la contraseña, código de acceso o clave de cifrado de un dispositivo electrónico? ¿Requerir que un sospechoso proporcione esta información viola las protecciones de autoincriminación de la Quinta Enmienda de EE. UU. de la persona? Las respuestas indicarán que las contraseñas, códigos de acceso y otros mecanismos de bloqueo de dispositivos electrónicos almacenados en la mente de una persona son más propensos a recibir protección bajo la Quinta Enmienda porque generalmente un acusado no tiene que compartir el contenido de su mente (Lyn, 2020, 427-428).

En este sentido, algunos controles más seguros vinculados a identificadores biométricos para bloquear y desbloquear dispositivos son menos propensos a recibir protección bajo la Quinta Enmienda; es decir, una persona podría ser obligada a proporcionar una muestra de sangre o un documento escrito porque no son actos testimoniales. Sin embargo, algunos tribunales podrían obligar a una persona a abrir su dispositivo electrónico protegido con un mecanismo de bloqueo biométrico, pero no revelar una contraseña de la mente (Lyn, 2020, 429).

En resumen, la protección de los datos mentales y las ideas es fundamental para los neuroderechos, que defienden la privacidad individual y el control sobre los pensamientos. Este debate plantea cuestiones legales y éticas sobre la revelación de información almacenada en la mente, como contraseñas y códigos de acceso, y las respuestas a estas preguntas afectarán la interpretación de

protecciones constitucionales y la formulación de políticas digitales que respeten la autonomía de los individuos.

Los datos derivados de neurotecnologías avanzadas enfrentan desafíos éticos y legales al vincularse con derechos fundamentales como la privacidad y la libertad, más allá de la protección de datos personales. Su regulación debe prevenir accesos no consentidos a pensamientos o memorias, como protege la Constitución chilena en cuanto a la actividad cerebral. Un enfoque cabal, debe equilibrar el avance tecnológico con la protección de derechos humanos, garantizando políticas que respeten la autonomía individual y prevengan abusos en el manejo de datos mentales.

2.1.3.2 Datos sintéticos

Ahora bien, hablando de otro tipo de dato, es necesario abordar los **datos sintéticos**. Los datos sintéticos son datos artificiales que se generan a partir de datos originales y un modelo que está entrenado para reproducir las características y la estructura de los datos originales (Riemann, s.f.). Se puede percibir como un enfoque de confidencialidad donde se comparten los datos sintéticos en lugar de los datos reales. Además de los métodos conocidos de desidentificación, los datos sintéticos son otro enfoque de despersonalización. Los datos sintéticos tienen las mismas propiedades estadísticas que los datos reales y pueden ser utilizados como un sustituto en sistemas de IA, pruebas de software y otros propósitos (WEF, 2020c).

Con los datos sintéticos, los valores de las variables sensibles se reemplazan con valores sintéticos generados por simulación. De alguna manera, los valores sintéticos son valores aleatorios generados por un simulador de función de distribución de probabilidad, por lo que estas distribuciones se seleccionan para reproducir tantas relaciones como sea posible en los datos originales (Mortazavi & Khaled, 2015, 52). Es importante destacar que estos datos se crean mezclando los elementos de un conjunto de datos o creando nuevos valores basados en los datos originales para que los totales y los valores

generales del conjunto se conservan, pero no están relacionados con ningún sujeto en particular (ICO, 2012). Es importante destacar que generar datos sintéticos a través del aprendizaje automático permite compartir datos con terceros sin poner en peligro la privacidad, empero, el sentido estricto, el procesamiento hacía datos sintéticos si se considera tratamiento de datos personales.

Los datos sintéticos, generados para replicar propiedades estadísticas de datos reales, permiten compartir información sin exponer datos sensibles. Aunque no están vinculados a individuos, su creación implica procesar datos personales originales, sometiéndolos a regulaciones de protección de datos. Para un uso responsable, es necesario generar estos valores preservando patrones y relaciones originales sin comprometer la privacidad, además de implementar mecanismos de validación para prevenir re-identificaciones accidentales o usos indebidos.

2.1.3.3 Datos degradados / desvanecidos y perecederos

Una vez agotada la reflexión sobre datos sintéticos, es necesario continuar con otros novísimos tipos de datos que no se han tenido en cuenta por las regulaciones; **Datos degradados o desvanecidos** (*Degrading or fading data*). Estos datos son una forma atractiva de proteger la privacidad de la información, ya que los datos personales son un activo valioso para las entidades. En la práctica, cuando se recopilan datos personales, la decisión del usuario sobre cuánto tiempo se conservarán los datos ni siquiera se tiene en cuenta, lo que lleva al controlador a retener los datos durante el mayor tiempo posible.

Dicho esto, la teoría de datos degradados surge para evitar riesgos de privacidad en el período de retención, permitiendo que los datos se retengan durante el menor tiempo posible, y aunque la retención sea limitada, se pretende crear una transición hacia la degradación de datos, lo que significa almacenar datos de manera progresivamente menos precisa, equilibrando así la privacidad y el uso de datos (Heerde et al., 2009, 146-153). Por ejemplo, la información sobre las coordenadas GPS exactas de un usuario podría intercambiarse por el nombre

de la calle, el barrio y la ciudad. Siguiendo el concepto anterior y con el objetivo de reducir el riesgo para la privacidad al publicar información espacial, algunas opciones serían aumentar el área de mapeo para abarcar más propiedades u ocupantes, reducir la frecuencia u oportunidad de publicación para que cubra más eventos, para que sea más difícil identificar un caso reciente o no revele datos adicionales como la hora o fecha del evento (ICO, 2012).

No obstante, si los datos se consideran datos personales o no, se determina principalmente por su contenido y su capacidad para relacionarse con individuos identificados o identificables, más que por su estado de degradación o desvanecimiento. Las organizaciones deben gestionar los datos personales de acuerdo con las leyes de protección de datos y las mejores prácticas para proteger los derechos de privacidad de las personas.

Un concepto que podría derivarse de la degradación o el desvanecimiento de datos es el de **datos perecederos**. Estos datos son información cuyo valor puede disminuir sustancialmente durante un tiempo especificado; se produce una disminución significativa en el valor cuando las circunstancias operativas que cambian hasta el punto de que la información ya no es útil (NIST, s.f.).

La diferencia entre los datos degradados y los datos perecederos radica en que los datos degradados se refieren a la disponibilidad e importancia de la información, mientras que los datos perecederos se refieren al valor. Aunque la visibilidad de la información (disponibilidad) podría permitir la reutilización de datos, hay que tener en cuenta que no todos los datos tienen el mismo valor, especialmente cuando los datos son incompletos, inconsistentes y atípicos para el mundo. Así que los datos valiosos solo se alcanzan cuando la información es explotada, pero no todos los datos son valiosos, por lo que debe haber más formas de acceder a ellos. Por lo tanto, los datos degradados pueden ser explotados, mientras que los datos perecederos son difíciles de usar porque necesitan contexto u otra información adicional.

Los datos perecederos generalmente se refieren a datos con una vida útil limitada que pierden su relevancia o utilidad con el tiempo. Si se considera que los datos perecederos son datos personales depende de la naturaleza de la

información que contienen y su potencial para identificar a individuos. En resumen, si los datos perecederos se consideran datos personales depende del contenido y contexto específicos. Los datos perecederos que contienen información personal están sujetos a regulaciones de protección de datos y deben ser manejados con cuidado para proteger los derechos de privacidad de las personas.

Un experto debe gestionar los datos degradados y perecederos estratégicamente, equilibrando utilidad y privacidad. Los datos degradados reducen progresivamente su precisión, como reemplazar coordenadas GPS exactas por ubicaciones generales, para minimizar riesgos de privacidad cumpliendo normativas. En cambio, los datos perecederos pierden relevancia con el tiempo, aunque sigan siendo accesibles. Mientras los degradados buscan proteger la privacidad, los perecederos simplemente pierden utilidad. Una gestión efectiva requiere identificar su ciclo de vida y definir políticas claras para su retención, anonimización o eliminación cuando caduquen o dejen de ser útiles.

2.1.3.4 Datos fragmentados y aislados

En otro orden de ideas, también surgen los **datos fragmentados**, que se basan en una arquitectura de bases de datos que divide los datos para que residan en fragmentos distribuidos en nodos de bases de datos separados. Los fragmentos son autónomos y únicos; no comparten datos ni recursos informáticos, pero juntos representan un conjunto de datos lógico (WEF, 2020b). Aunque la razón principal para fragmentar los datos es porque el tamaño del estado es demasiado grande para ser servido por una sola máquina (Burns, 2018, 59), vale la pena mencionar que este principio de fragmentación podría llevar a técnicas de privacidad mejoradas, como el análisis federado, donde las partes comparten las ideas obtenidas del análisis de sus datos sin compartir los datos en sí, o la computación segura entre múltiples partes, donde el análisis de datos se distribuye entre varias partes de modo que ninguna de ellas pueda ver el conjunto completo de entradas (WEF, 2019d).

Transitando a otro tema, en la práctica también existen los **datos aislados o datos en silos (*siloed data*)**, que se utilizan para describir las diferentes partes de una organización que tienen datos almacenados de manera particular y que no se comparten fácilmente con otras partes de la organización (ya sea por razones técnicas, administrativas o políticas). Por lo tanto, cada entidad, departamento o servicio mantiene bases de datos separadas como un activo de cada división, pero los datos no funcionan bien en un entorno de datos aislados porque ayudan a descubrir patrones, problemas y oportunidades inesperadas (Thomas & Zikopoulos, 2020, 50).

Las razones conscientes para aislar los datos pueden ser preocupaciones técnicas (diferentes estándares tecnológicos, implementaciones o sistemas) y también pueden deberse a cuestiones de política y gobernanza (diferentes políticas legales, regulatorias, institucionales y normativas) (WEF, 2019e). Un ejemplo de este tipo de datos son los secretos comerciales. Por ejemplo, la fórmula de Coca-Cola se ha mantenido en secreto durante más de 130 años porque implementan muchos tipos diferentes de políticas para mantener la confidencialidad de los secretos comerciales, como el uso de mecanismos de control de acceso en sistemas informáticos, limitando el acceso físico según principios de necesidad de saber, e incluso utilizando principios de separación de funciones para asegurarse de que ninguna persona conozca un secreto comercial completo (Lyn, 2020, 259-260).

Los datos fragmentados y aislados presentan enfoques distintos en la gestión de información, cada uno con ventajas y desafíos. Los datos fragmentados se dividen en fragmentos autónomos distribuidos en varios nodos, ideales para manejar grandes volúmenes de información, garantizar escalabilidad y aplicar técnicas de privacidad como el análisis federado. Son útiles cuando la privacidad y la escalabilidad son esenciales. En contraste, los datos aislados se encuentran compartimentados dentro de una organización, con acceso limitado por razones técnicas, políticas o de confidencialidad, como en el caso de secretos comerciales. Sin embargo, este aislamiento puede restringir análisis integrales y descubrimiento de patrones. Para gestionar ambos tipos de datos, las organizaciones deben

equilibrar privacidad, confidencialidad y funcionalidad: garantizar interoperabilidad y seguridad en los fragmentados, y fomentar la integración selectiva en los aislados, respetando restricciones legales y de gobernanza.

Capítulo 3
ACTORES Y SUS ROLES EN LOS DATOS.

Capítulo 3. Actores y sus Roles en los Datos

Una vez que se han explorado las teorías sobre los datos y sus clasificaciones, el siguiente paso consiste en analizar a los actores y sus funciones dentro del ecosistema de los datos.

Sin más preámbulo, en la gestión de datos, donde la información es un activo crucial y persistente, los actores que intervienen en su ciclo de vida desempeñan roles distintos pero interconectados que dan forma a cómo se manejan, protegen y utilizan los datos. Desde el individuo cuyos datos personales son el núcleo de la importancia, hasta los líderes estratégicos en las organizaciones que determinan políticas y procesos; cada actor tiene un papel esencial que desempeñar en el ecosistema digital.

En este sentido, el presente Capítulo aborda la diversidad de actores y sus roles en relación con los datos, destacando cómo cada uno contribuye a la creación, flujo y aprovechamiento de información en diferentes contextos y sectores. Desde los interesados, cuyos derechos y expectativas deben ser salvaguardados, hasta los profesionales especializados encargados de la gestión técnica y estratégica de los datos, se examinarán detenidamente cómo estas figuras interactúan y colaboran para asegurar la confidencialidad, integridad y disponibilidad de los datos, así como para maximizar su valor y utilidad.

En este apartado se analizarán desde los roles fundamentales como el Interesado (*data subject*), aquel cuyos datos son objeto de tratamiento, hasta las figuras más especializadas como el Científico de Datos (*Data Scientist*) o el Arquitecto de Datos (*Data Architect*), cuyas habilidades y conocimientos técnicos son esenciales para desbloquear el potencial de los datos. Además, se abordarán los roles institucionales y de liderazgo, como los Oficiales/Directores vinculados a los datos (*C-level*) y los comités de gobernanza, que establecen la visión estratégica y los marcos regulatorios para la gestión responsable y ética de la información.

Finalmente, se expondrá cómo cada actor desempeña un rol único pero interdependiente en la gestión y gobernanza de los datos, lo que subraya la

importancia de una colaboración efectiva y una comprensión clara de las responsabilidades y expectativas en este entorno digital en constante evolución. Desde la protección de la privacidad individual hasta la optimización de los procesos empresariales, cada actor y su rol en los datos son elementos vitales en la economía digital basada en datos.

Figura 4. Actores en el tratamiento de datos.

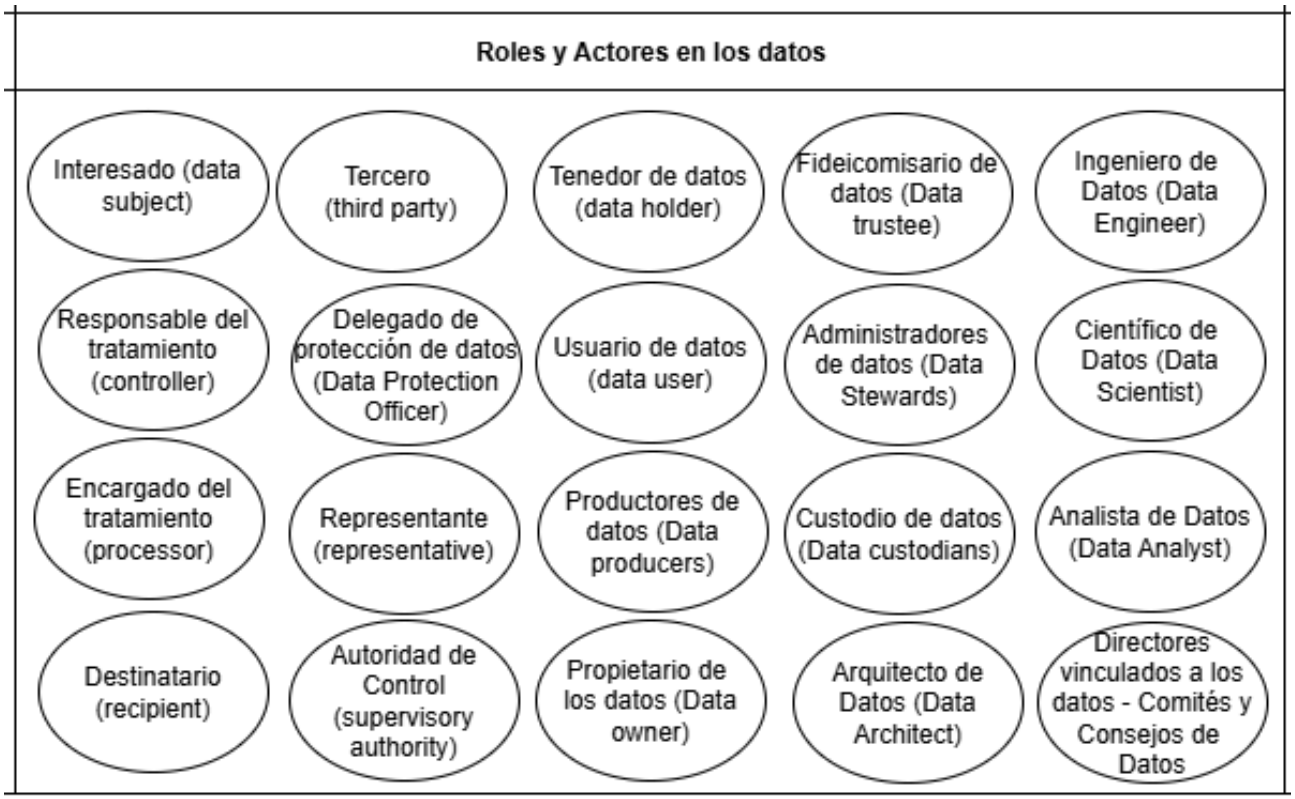

Figura de elaboración propia.

3.1 Interesado (data subject)

El término "interesado" o "*data subject*" se refiere a cualquier individuo cuyos datos personales son procesados por una organización. Este concepto es clave en la protección de datos, ya que representa a la persona sobre la cual se manejan datos personales.

De manera más formal, de acuerdo con el artículo 4 1) del RGPD menciona que (2016):

> «datos personales»: toda información sobre una persona física identificada o identificable («el interesado»); se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona.

El término interesado, también se asocia al concepto *data subject*. Cabe destacar que el Considerando 14 del RGPD específica en lo general que el Reglamento se aplica a individuos, sin importar su nacionalidad o ubicación, en lo que respecta al tratamiento de sus datos personales. Sin embargo, no se aplica al tratamiento de datos de personas jurídicas, como empresas, incluyendo su nombre, forma legal y detalles de contacto. Y en el mismo orden de ideas, el Considerando 27, complementa lo dicho al estipular que el Reglamento no cubre la protección de datos personales de personas fallecidas y que cada Estado miembro tiene la autoridad para establecer normativas sobre el manejo de estos datos.

En la UE, algunos países como Bulgaria, Hungría, Italia, Portugal, Eslovaquia y España otorgan derechos específicos a herederos o familiares de personas fallecidas sobre el manejo de sus datos personales: en Hungría y España, los parientes cercanos o ejecutores designados pueden ejercer los derechos del difunto; en Italia, un agente designado puede hacerlo; además, la ley francesa permite a los sujetos de datos dar instrucciones sobre sus datos después de su muerte (Tang, 2023, 11).

Cabe destacar que en algunas jurisdicciones como en la mexicana, el *Interesado* es equivalente a la figura denominada *Tienedor de los datos personales*, pero esta definición entraría en conflicto con los nuevos roles que toma la UE en sus nuevas normativas de Reglamento de Datos y Reglamento de Gobernanza de Datos. Por lo tanto, por cuestiones de homologación y tomar en

cuenta el mejor estándar actual, se seguirá la visión de la UE con el termino de interesada o data subject.

Adicionalmente, en el caso de México, el tenedor de los datos personales es la persona física a quien corresponden o conciernen los datos personales sujetos a tratamiento, siendo considerado sujeto de protección del derecho a la protección de datos. Tanto la Ley Federal de Protección de Datos Personales en Posesión de los Particulares como la Ley General de Protección de Datos personales en Posesión de Sujetos Obligados definen en idénticos términos la figura del tenedor de los datos. El interesado no es el dueño pero si el portador de los datos personales, aunque estos estén en posesión de un tercero para su tratamiento; es importante destacar que el titular es una persona física con capacidad jurídica, siendo el sujeto de protección de la normatividad de datos personales y el titular del derecho humano a la protección de datos personales (Davara et al., 2019, 850-853).

En conclusión, el papel del interesado, o *data subject*, es crucial en el contexto de la protección de datos, toda vez que representa los derechos y las preocupaciones de los individuos cuyos datos personales son procesados por organizaciones. La protección y el respeto de los derechos del interesado son fundamentales para garantizar la transparencia, la equidad y la seguridad en el manejo de los datos personales. Por lo tanto, es esencial que las organizaciones comprendan y respeten los derechos del interesado, promoviendo la confianza y el cumplimiento de la regulación en protección de datos.

Por ejemplo, una persona que sea un interesado o *data subject* debe estar bien informado sobre sus derechos y responsabilidades en relación con sus datos personales. En el caso de que una persona recibe una notificación de su banco indicando que sus datos serán procesados para fines de marketing. Como interesado, esta persona podría ejercer su derecho de acceso al solicitar al banco detalles sobre qué datos personales están siendo procesados, con qué propósito, durante cuánto tiempo serán conservados y si serán compartidos con terceros. Si considera que el procesamiento no es adecuado o excede lo autorizado, puede

ejercer su derecho de oposición o, en su caso, solicitar la rectificación o eliminación de sus datos.

3.2 Responsable del tratamiento (controller)

El término "Responsable del Tratamiento" o "*controller*" se refiere a la entidad que determina los fines y los medios del procesamiento de datos personales. Esencialmente, el *controller* es responsable de asegurar que el procesamiento de datos se realice de manera conforme a las leyes de privacidad y protección de datos.

Formalmente, el Responsable del tratamiento, o también conocido como *controller*, de acuerdo al artículo 4 7) del RGPD (2016):

> «responsable del tratamiento» o «responsable»: la persona física o jurídica, autoridad pública, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento; si el Derecho de la Unión o de los Estados miembros determina los fines y medios del tratamiento, el responsable del tratamiento o los criterios específicos para su nombramiento podrá establecerlos el Derecho de la Unión o de los Estados miembros;

Los controladores/responsables de datos son responsables de garantizar que los datos personales se procesen de acuerdo con el Reglamento. El controlador de datos es la parte que determina los propósitos y medios del procesamiento de datos personales, ya sea de manera individual o conjunta. Esto incluye decidir qué datos se recopilarán, de quién se recopilarán, si hay justificación para no notificar a los sujetos de datos o solicitar su consentimiento, cuánto tiempo se retendrán los datos, entre otros aspectos. Además, es responsabilidad del controlador de datos garantizar que cualquier procesador cumpla con las reglas establecidas en el Reglamento. En caso de que haya múltiples controladores, es fundamental establecer claramente las responsabilidades de cumplimiento con el Reglamento antes de realizar cualquier

procesamiento o recopilación de datos personales (IT Governance Privacy Team, 2020, 27-30).

Las responsabilidades del controlador de datos incluyen garantizar el cumplimiento de una solicitud de acceso por parte del sujeto, diseñando un proceso que cumpla con los requisitos específicos del RGPD y capacitando al personal esencial. Es importante establecer acuerdos de nivel de servicio que garanticen la respuesta oportuna de los procesadores de datos. Además, el controlador debe proporcionar al sujeto datos, información sobre sus derechos de acceso, rectificación, cancelación u oposición al procesamiento, así como el derecho a presentar una queja ante la autoridad supervisora. Es esencial identificar el origen de los datos personales recopilados y facilitar detalles sobre las salvaguardias aplicadas en caso de transferencias fuera de la UE u organizaciones internacionales. La transparencia es fundamental en cada etapa de una solicitud de acceso (IT Governance Privacy Team, 2020, 147-148).

En conclusión, el Responsable del Tratamiento, o *controller*, es una figura central en el ámbito de la protección de datos y la privacidad. Su papel radica en garantizar que el procesamiento de datos personales se lleve a cabo de manera legal, ética y segura, cumpliendo con las normativas de privacidad y protección de datos correspondientes. Al asumir la responsabilidad de determinar los propósitos y los medios del tratamiento de datos, el *controller* se convierte en el garante de los derechos de los interesados y en el protector de su privacidad. Su gestión adecuada contribuye a fomentar la confianza de los usuarios y a mantener la integridad y la seguridad de la información personal en el entorno digital actual.

El responsable del tratamiento debe asegurar que todas las actividades de manejo de datos personales cumplan con las normativas aplicables. Por ejemplo, al recopilar datos para ofrecer descuentos personalizados, debe informar a los clientes sobre los datos recopilados, su propósito, tiempo de conservación y posibles destinatarios. Además, debe implementar medidas como cifrado y controles de acceso para proteger la información. Si un cliente ejerce su derecho de acceso, el responsable debe responder de forma completa y oportuna, garantizando transparencia y respeto a los derechos del interesado.

3.3 Encargado del tratamiento (processor)

El término "Encargado del Tratamiento" o "*processor*" se refiere a la entidad que procesa datos personales en nombre del Responsable del Tratamiento, de conformidad con sus instrucciones. El Encargado del Tratamiento desempeña un papel importante en el cumplimiento de las obligaciones legales en materia de protección de datos, toda vez que actúa como una entidad de confianza para el Responsable del Tratamiento al manejar datos en su nombre.

Formalmente, de acuerdo con el artículo 4 8) del RGPD (2016), se entiende por "«encargado del tratamiento» o «encargado»: la persona física o jurídica, autoridad pública, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento".

El concepto de procesador implica una entidad que participa en el procesamiento de datos personales bajo las directrices del controlador. La definición de procesador según la regulación implica dos elementos clave: que la persona sea una entidad legal separada respecto al controlador y que procese datos personales en nombre del controlador. Aunque puede tener un papel más activo en el proceso, el controlador retiene la responsabilidad principal. El procesador debe seguir las instrucciones del controlador y mantener la confidencialidad, seguridad y registros adecuados. Si el procesador toma decisiones sobre el propósito o los medios del procesamiento más allá de su mandato, puede considerarse un controlador en ese aspecto, lo que aumenta su responsabilidad. La regulación requiere que el procesador opere solo bajo las instrucciones del controlador y establezca un contrato por escrito que defina claramente las responsabilidades y los términos del procesamiento de datos. Además, se establecen requisitos para la subcontratación de procesadores y se proporcionan criterios para distinguir entre los roles de controlador y procesador, como el nivel de instrucción previa y el monitoreo de la ejecución del servicio (Macmillan, 2018, 90-92).

Aunque diversas leyes y regulaciones puedan tener definiciones ligeramente diferentes para el controlador de datos y el procesador de datos, los principios fundamentales para definir los roles son bastante similares. En general, un controlador de datos es una persona natural o jurídica, autoridad pública, agencia u otro organismo que, solo o junto con otros, determina los fines y medios del procesamiento. Un procesador de datos es una persona natural o jurídica, autoridad pública, agencia u otro organismo que procesa datos personales en nombre de un controlador.

Dado que el controlador de datos y el procesador de datos tienen diferentes obligaciones bajo las regulaciones de protección de datos, una organización debe definir su rol de procesamiento de datos personales en función de si la organización determina los fines y medios de las actividades de procesamiento de datos. Además, en algunos casos, el manejo de información personal puede necesitar asumir responsabilidades adicionales; es importante tener en cuenta que la determinación del "propósito" del procesamiento está reservada al "controlador", mientras que la determinación de los "medios" del procesamiento puede delegarse por el controlador. Por lo tanto, un controlador de datos puede ser a la vez un controlador y un procesador de datos al mismo tiempo para diferentes actividades de procesamiento, aunque no puede ser controlador y procesador en relación con las mismas actividades de procesamiento (Tang, 2023, 116-118).

En conclusión, el Encargado del Tratamiento, o procesador, desempeña un papel esencial en el cumplimiento de las obligaciones legales en materia de protección de datos al procesar información personal en nombre del Responsable del Tratamiento. Este rol se define claramente en el RGPD así como en otras normatividades de protección de datos, donde se establece que el procesador opera bajo las instrucciones del controlador y debe mantener la confidencialidad, seguridad y registros adecuados. Aunque el procesador puede tener un papel activo en el proceso, el controlador retiene la responsabilidad principal, y ambos deben cumplir con los requisitos legales y contractuales para garantizar la protección adecuada de los datos personales. Es determinante comprender la

distinción entre estos roles y sus responsabilidades, ya que esto determina cómo se gestionan y protegen los datos personales según las regulaciones aplicables.

El encargado del tratamiento debe seguir las instrucciones del responsable y garantizar medidas de seguridad adecuadas para los datos personales procesados. Por ejemplo, una empresa de servicios en la nube contratada por un hospital debe almacenar los datos médicos de manera segura, utilizando cifrado y controles de acceso, y procesarlos exclusivamente para los fines definidos en el contrato, como gestionar historiales médicos. Si subcontrata servicios, debe obtener autorización previa y asegurar que el subcontratista cumpla las mismas normas de seguridad y confidencialidad. También debe mantener un registro de actividades de tratamiento y notificar al hospital cualquier incidente de seguridad de forma inmediata.

3.4 Destinatario (recipient)

De acuerdo con el artículo 4 9) del RGPD (2016) se entiende por destinatario o *recipient*:

> La persona física o jurídica, autoridad pública, servicio u otro organismo al que se comuniquen datos personales, se trate o no de un tercero. No obstante, no se considerarán destinatarios las autoridades públicas que puedan recibir datos personales en el marco de una investigación concreta de conformidad con el Derecho de la Unión o de los Estados miembros; el tratamiento de tales datos por dichas autoridades públicas será conforme con las normas en materia de protección de datos aplicables a los fines del tratamiento.

No obstante, en el artículo 2 14) del Reglamento de Datos de la UE (2023), define un rol de destinatario, pero sin vincular al RGPD como lo hace con otros conceptos como el de datos personales. El nuevo Reglamento de datos define el destinatario de datos o *data recipient* como:

> Una persona física o jurídica que actúa con un propósito relacionado con su actividad comercial, empresa, oficio o profesión, distinta del usuario de un producto

> conectado o servicio relacionado, a disposición de la cual el titular de datos pone los datos, incluso un tercero previa solicitud del usuario al titular de datos o de conformidad con una obligación legal en virtud del Derecho de la Unión o de la normativa nacional adoptada de conformidad con el Derecho de la Unión;

En una óptica muy general el artículo 4 9) del RGPD (2016) define al destinatario o receptor como cualquier persona física o jurídica, autoridad pública, servicio u otro organismo al que se comuniquen datos personales, ya sea que se trate de un tercero o no. No obstante, excluye de esta definición a las autoridades públicas que puedan recibir datos personales en el marco de una investigación concreta, siempre que el tratamiento de esos datos cumpla con las normas de protección de datos aplicables.

En contraste, el artículo 2(14) del Reglamento de Datos (2023) define al destinatario de datos como una persona física o jurídica que actúa con un propósito relacionado con su actividad comercial, empresa, oficio o profesión, distinta del usuario de un producto o servicio relacionado. Este Reglamento establece condiciones específicas para la puesta a disposición de datos por parte de los titulares de datos a los destinatarios, incluyendo la necesidad de acordar modalidades de puesta a disposición, condiciones justas y no discriminatorias, entre otros aspectos.

Aunque el concepto de destinatario se aborda tanto en el RGPD como en el Reglamento de Datos, existen diferencias en sus definiciones y alcances. Mientras que el RGPD proporciona una definición más amplia y general del destinatario, el Reglamento de Datos establece una definición más específica, relacionada con actividades comerciales y profesionales. Ambos reglamentos comparten el objetivo común de garantizar la protección de los datos personales y establecer condiciones justas y transparentes para su tratamiento y comunicación.

Al respecto, los EDPB y EDPS de la UE en una opinión conjunta, declararon en su momento, previo a la aprobación del Reglamento de Datos que, la propuesta introducía nuevas definiciones y hacía referencia a conceptos específicos en otras regulaciones, a pesar de que el RGPD ya proporciona algunas definiciones relevantes; destacaron que la introducción de la definición de

'destinatario de datos', ya estaba definida en el RGPD y por lo tanto, sugirieron aclarar la necesidad de estas nuevas definiciones o identificar las del RGPD que no se aplican (EDPB-EDPS, 2022a, 12-14).

Y, por otra parte, también estas autoridades reconocieron que el ámbito de aplicación previsto de la propuesta no se refiere exclusivamente a datos personales, sino que se aplicaría tanto a datos personales como no personales generados por el uso de productos o servicios según el significado de la propuesta. Además, destacaron que el derecho a la protección de datos personales es inalienable y no puede renunciarse mediante un acuerdo entre el tenedor de datos y el destinatario de datos, y se instó a especificar claramente las obligaciones de los tenedores y destinatarios de datos en cumplimiento con el RGPD para garantizar una mayor protección de los datos personales (EDPB-EDPS, 2022b, 8-19).

En resumen, el término "Destinatario" o "*recipient*" se define de manera amplia en el RGPD (2016), abarcando a cualquier persona física, entidad u organización que reciba o tenga acceso a datos personales, ya sea que se trate de un tercero o no. Sin embargo, excluye a las autoridades públicas que puedan recibir datos en el contexto de una investigación específica, siempre que el tratamiento cumpla con las normativas de protección de datos. Por otro lado, el Reglamento de Datos de la UE (2023) proporciona una definición más específica, vinculada a actividades comerciales y profesionales, donde el destinatario es aquel que actúa con un propósito relacionado con su actividad, distinto al usuario de un producto o servicio relacionado. Aunque ambos reglamentos comparten el objetivo de proteger los datos personales, existen diferencias en sus enfoques y definiciones, lo que ha llevado a debates sobre la necesidad y claridad de estas distinciones en las regulaciones.

Por ejemplo, un destinatario, como una empresa de análisis de datos que recibe información de un minorista para estudiar tendencias de consumo, debe garantizar que los datos sean utilizados únicamente para el propósito acordado, respetando las normativas como el RGPD, mediante medidas de seguridad como el cifrado o la seudonimización. Además, debe evitar la divulgación no autorizada

de los datos y, al finalizar el contrato, asegurarse de eliminarlos o devolverlos según lo establecido, protegiendo siempre los derechos del interesado y cumpliendo con las obligaciones legales y contractuales.

3.5 Tercero (third party)

De acuerdo con el artículo 4 10) del RGPD (2016), un tercero o *third party*, es una "persona física o jurídica, autoridad pública, servicio u organismo distinto del interesado, del responsable del tratamiento, del encargado del tratamiento y de las personas autorizadas para tratar los datos personales bajo la autoridad directa del responsable o del encargado".

A *prima facie*, los conceptos de destinatarios y tercero son complejo de identificar, incluso para expertos en protección de datos. No obstante, el punto medular podría radicar en el punto de revelación/divulgación en contraposición con el de transferencia de la información.

Es decir, en el contexto del RGPD, un destinatario se refiere a una persona física o jurídica, autoridad pública, agencia u otro organismo al que se revelan datos personales, ya sea que sea un tercero o no. Por otro lado, un tercero se refiere a una persona física o jurídica, autoridad pública, agencia u otro organismo distinto del sujeto de los datos, el responsable del tratamiento, el encargado del tratamiento y las personas que, bajo la autoridad directa del responsable del tratamiento o del encargado del tratamiento, estén autorizadas a procesar datos personales.

La diferencia principal radica en que el destinatario puede ser cualquier entidad que reciba datos personales, ya sea dentro o fuera de la organización que los procesa, mientras que un tercero es específicamente cualquier entidad externa a la relación directa entre el sujeto de los datos y el responsable o encargado del tratamiento.

En lenguaje más sencillo y de manera más directa, en el suponer que una empresa envía los datos personales de sus clientes a un departamento interno, como el equipo de ventas, para procesar pedidos y proporcionar atención al cliente. En este caso, el equipo de ventas es el destinatario de los datos personales proporcionados por la empresa. Por otra parte, si esa misma empresa utiliza un servicio externo de correo electrónico para enviar boletines informativos a sus clientes. El proveedor de servicios de correo electrónico es un tercero, ya que no es parte de la organización interna de la empresa, pero aun así maneja los datos personales de los clientes en nombre de la empresa.

En resumen, el RGPD (2016) define tanto los destinatarios como los terceros en el contexto del tratamiento de datos personales. Un destinatario se refiere a cualquier entidad que recibe datos personales, ya sea dentro o fuera de la organización que los procesa, mientras que un tercero es específicamente cualquier entidad externa a la relación directa entre el sujeto de los datos y el responsable o encargado del tratamiento. La diferencia clave radica en el punto de revelación o divulgación de los datos. Mientras que un destinatario recibe datos personales, un tercero es una entidad externa que no forma parte directa de la relación de tratamiento de datos, pero aun así maneja datos personales en nombre del responsable del tratamiento.

Por ejemplo, un tercero, como una empresa de servicios de almacenamiento en la nube contratada por un hospital para guardar historiales médicos, debe garantizar que el tratamiento de los datos personales cumpla estrictamente con las instrucciones del responsable del tratamiento y las normativas aplicables, como el RGPD. Aunque no tiene una relación directa con el interesado, debe implementar medidas de seguridad adecuadas, mantener la confidencialidad y evitar usar los datos para fines propios, asegurando así la protección de los derechos de los titulares de los datos.

3.6 Delegado de protección de datos (Data Protection Officer)

La Sección 4 del RGPD (2016) aborda la designación y funciones del Delegado de Protección de Datos o *Data Protection Officer*. Específicamente, el artículo 37 establece los criterios para su designación, que incluyen el tipo de tratamiento realizado (especialmente si involucra datos sensibles o a gran escala) y la naturaleza de la entidad responsable. Se permite que un grupo empresarial nombre un único delegado de protección de datos accesible desde cada establecimiento, mientras que las autoridades públicas pueden designar uno para varias entidades considerando su estructura y tamaño.

En el mismo orden de ideas, el artículo 38 describe la posición del Delegado de Protección de Datos, enfatizando su independencia y la obligación del responsable y encargado de respaldarlo. Los interesados pueden comunicarse directamente con el Delegado, quien debe mantener la confidencialidad y no recibir instrucciones en sus funciones. El artículo 39 enumera las responsabilidades mínimas del Delegado, que incluyen asesorar sobre el cumplimiento normativo, supervisar operaciones de tratamiento, cooperar con autoridades y actuar como punto de contacto para cuestiones de protección de datos. El Delegado debe considerar los riesgos asociados al tratamiento en el desempeño de sus funciones.

Cabe destacar que, el representante y el delegado de protección de datos no son lo mismo. Tanto el representante como el delegado de protección de datos son figuras vitales para asegurar el cumplimiento normativo. El representante se designa cuando un responsable o encargado de tratamiento opera fuera de la Unión Europea (UE), actuando como su punto de contacto para las autoridades y los interesados dentro de la UE. Por otro lado, el Delegado es nombrado por un responsable o encargado de tratamiento dentro de la UE y tiene la responsabilidad de supervisar el cumplimiento del RGPD, asesorar internamente sobre temas de protección de datos y actuar como enlace con la autoridad de control y los interesados. Mientras que el representante se enfoca en la representación externa, el Delegado se concentra en las operaciones internas y el cumplimiento normativo.

Por otra parte, el delegado de protección de datos y un **Oficial de privacidad (Chief Privacy Officer)** o director tampoco es lo mismo. La diferencia entre un Delegado de Protección de datos y un Director u oficial de privacidad dentro de una organización radica en su alcance y función. El Delegado del RGPD, es designado según los requisitos establecidos por la normatividad y tiene la responsabilidad de supervisar el cumplimiento del RGPD, asesorar internamente sobre cuestiones de protección de datos y actuar como enlace con la autoridad de control y los interesados. En cambio, el oficial o director de privacidad dentro de una organización es un puesto interno creado por la propia empresa, con un enfoque más amplio que puede abarcar no solo el cumplimiento del RGPD, sino también otras leyes y regulaciones o estándares de privacidad de datos importantes para la organización, además de ser responsable de desarrollar y mantener políticas internas de privacidad de datos y gestionar incidentes de seguridad de datos. Mientras que la designación del Delegado del RGPD está sujeta a requisitos legales específicos, el Director u oficial de privacidad dentro de una organización es una decisión interna de la empresa para fortalecer su gestión de la privacidad de datos.

En conclusión, la Sección 4 del RGPD (2016) establece las directrices para la designación y funciones del Delegado de Protección de Datos (DPO). Este profesional es esencial para garantizar el cumplimiento normativo en materia de protección de datos y actúa como un punto focal independiente dentro de la organización. El DPO debe ser designado en función de criterios específicos, como el tipo de tratamiento realizado y la naturaleza de la entidad responsable. Además, el RGPD enfatiza la independencia del DPO y enumera sus responsabilidades mínimas, que incluyen asesorar sobre el cumplimiento normativo, supervisar operaciones de tratamiento y cooperar con autoridades y actuar como punto de contacto para cuestiones de protección de datos. Es importante distinguir entre el DPO y otras figuras dentro de la organización, como el representante y el Oficial de Privacidad, ya que cada uno tiene roles y responsabilidades distintos en el cumplimiento de las regulaciones de protección de datos.

Un DPO designado por un hospital que gestiona historiales médicos a gran escala debe garantizar que las operaciones de tratamiento cumplan con el RGPD, asesorando al personal sobre el manejo adecuado de datos sensibles, supervisando los procesos de anonimización y pseudonimización, y asegurando que las medidas de seguridad sean robustas. Además, debe actuar como punto de contacto entre el hospital y la autoridad de protección de datos, así como atender consultas de los pacientes relacionados con sus derechos, todo mientras mantiene su independencia y confidencialidad.

3.7 Representante (representative)

De acuerdo con el artículo 4 17) del RGPD (2016), se define representante o *representative*, como una:

> "persona física o jurídica establecida en la Unión que, habiendo sido designada por escrito por el responsable o el encargado del tratamiento con arreglo al artículo 27, represente al responsable o al encargado en lo que respecta a sus respectivas obligaciones en virtud del presente Reglamento",

Cabe destacar que el artículo 27 que se menciona en la definición de representante, establece la obligación para los responsables o encargados del tratamiento no establecidos en la Unión Europea de designar un representante en un Estado miembro donde se encuentren los interesados cuyos datos se traten en el contexto de ofrecer bienes o servicios, o cuyo comportamiento se esté monitoreando. Esta obligación no aplica a tratamientos ocasionales que no involucren datos sensibles ni a autoridades u organismos públicos. En cumplimiento del RGPD, el representante designado será responsable de atender consultas de autoridades de control y de interesados, sin perjuicio de las acciones que puedan tomarse contra el responsable o encargado directamente.

Este representante, designado por escrito, actúa como enlace entre la organización y las autoridades de supervisión y los sujetos de datos, asegurando el cumplimiento normativo y facilitando la comunicación. El incumplimiento de

estas obligaciones puede resultar en multas considerables, como se vio en el caso del Tribunal Regional de La Haya contra WhatsApp, donde se ordenó a la empresa designar un representante en los Países Bajos para evitar sanciones económicas (Tang, 2023, 129-131).

Cabe destacar que este representante, que puede ser una persona física o jurídica, tiene un papel mayormente pasivo, siendo identificado en avisos de privacidad, y puede ser contactado por autoridades de protección de datos y sujetos de datos en asuntos relacionados con el procesamiento de datos. Además, el representante debe mantener registros de actividades de procesamiento y cooperar con las autoridades de protección de datos a pedido. Las empresas multinacionales deben considerar designar una filial en un Estado miembro de la UE favorable para sus operaciones, lo que podría facilitar el tratamiento simplificado de la RGPD y la jurisdicción única de una autoridad de protección de datos de la UE (Determann, 2022, 11-12).

En resumen, el RGPD (2016) define al representante como una persona física o jurídica establecida en la Unión Europea (UE) y designada por escrito por el responsable o encargado del tratamiento, según lo establecido en el artículo 27. Este representante actúa como enlace entre la organización y las autoridades de supervisión y los sujetos de datos, asegurando el cumplimiento normativo y facilitando la comunicación. La designación de un representante es obligatoria para los responsables o encargados del tratamiento no establecidos en la UE, si están sujetos a ciertas condiciones, como el ofrecimiento de bienes o servicios a personas en la UE o el monitoreo de su comportamiento. El incumplimiento de estas obligaciones puede resultar en sanciones económicas. El representante, aunque tiene un papel mayormente pasivo, debe mantener registros de actividades de procesamiento y cooperar con las autoridades de protección de datos a pedido. Las empresas multinacionales deben considerar cuidadosamente la designación de un representante en un Estado miembro de la UE para facilitar el cumplimiento del RGPD y evitar sanciones.

Por ejemplo, un representante designado por una empresa de EE. UU. que ofrece servicios en línea a clientes de la UE debe estar establecido en un Estado

miembro de la UE. Su función incluye recibir y responder consultas de autoridades de protección de datos y usuarios sobre el tratamiento de datos personales. Además, debe garantizar que la empresa mantenga registros de las actividades de procesamiento y coopere con las solicitudes de las autoridades, todo mientras facilita la comunicación entre las partes y asegura el cumplimiento del RGPD, actuando como el vínculo oficial entre la empresa y la Unión Europea.

3.8 Autoridad de Control (supervisory authority)

De acuerdo con el artículo 4 21) y 22), una autoridad de control o *supervisory authority*, es "la autoridad pública independiente establecida por un Estado miembro con arreglo a lo dispuesto en el artículo 51". Y en el mismo orden de ideas, una autoridad de control interesada o *supervisory authority concerned*:

> "la autoridad de control a la que afecta el tratamiento de datos personales debido a que:
> a) el responsable o el encargado del tratamiento está establecido en el territorio del Estado miembro de esa autoridad de control;
> b) los interesados que residen en el Estado miembro de esa autoridad de control se ven sustancialmente afectados o es probable que se vean sustancialmente afectados por el tratamiento, o
> c) se ha presentado una reclamación ante esa autoridad de control".

El Capítulo VI del RGPD (2016) establece las disposiciones sobre las autoridades de control independientes. En la Sección 1, se detalla la independencia de estas autoridades, destacando su responsabilidad en supervisar la aplicación del RGPD para proteger los derechos de las personas y facilitar la circulación de datos en la Unión Europea. Además, se establece la cooperación entre autoridades de control y se detallan las condiciones de su designación, garantizando su independencia y recursos adecuados.

En la Sección 2 el mismo capítulo del RGPD, se especifican las competencias, funciones y poderes de estas autoridades, que incluyen la promoción de la sensibilización sobre la protección de datos, el asesoramiento a instituciones y ciudadanos, la recepción de reclamaciones, la realización de investigaciones y la imposición de sanciones en caso de incumplimientos del RGPD. Asimismo, se exige la elaboración de informes anuales de actividad para garantizar la transparencia y rendición de cuentas.

En conclusión, el RGPD (2016) establece disposiciones claras sobre las autoridades de control, definidas como autoridades públicas independientes establecidas por los Estados miembros para supervisar la aplicación del RGPD. Estas autoridades juegan un papel importante en la protección de los derechos de las personas y en facilitar la circulación de datos en la UE. El capítulo VI del RGPD detalla la independencia de estas autoridades, así como sus competencias, funciones y poderes, que incluyen la promoción de la sensibilización sobre la protección de datos, el asesoramiento a instituciones y ciudadanos, la recepción de reclamaciones, la realización de investigaciones y la imposición de sanciones en caso de incumplimientos del RGPD. Además, se establece la cooperación entre autoridades de control y se garantiza su independencia y recursos adecuados para cumplir con sus responsabilidades de manera efectiva.

Por ejemplo, una Autoridad de Control, como la Agencia Española de Protección de Datos (AEPD) en España, supervisa la correcta aplicación del RGPD en su territorio. Entre sus funciones, atiende reclamaciones de los ciudadanos sobre el tratamiento de sus datos personales, realiza investigaciones en casos de posibles incumplimientos, y promueve la concienciación pública sobre la importancia de la protección de datos. Además, tiene potestad para imponer sanciones, como multas administrativas, en caso de infracciones graves, y trabaja en coordinación con otras autoridades de control de la UE para garantizar una aplicación coherente del RGPD en todos los Estados miembros.

3.9 Tenedor de datos (data holder)

Dejando por un lado el RGPD (2016) y asuntos específicos en materia de protección de datos, pasando a una visión de economía de datos, surge un nuevo rol dentro de las novísimas normativas relacionadas con los datos; el tenedor de datos. Dicho lo anterior, de acuerdo con el Reglamento de Datos (2023) y el Reglamento de Gobernanza de Datos (2022), un tenedor de datos o *data holder* es:

- Reglamento de Datos (2023): “una persona física o jurídica que tiene el derecho o la obligación, con arreglo al presente Reglamento, al Derecho de la Unión aplicable o a la normativa nacional adoptada de conformidad con el Derecho de la Unión, de utilizar y poner a disposición datos, incluidos, cuando se haya pactado contractualmente, los datos del producto o los datos de servicios relacionados que haya extraído o generado durante la prestación de un servicio relacionado”
- Reglamento de Gobernanza de Datos (2022): una persona jurídica o un interesado que, de conformidad con la legislación nacional o de la Unión, tiene derecho a conceder acceso a determinados datos personales o no personales que estén bajo su control, o a compartir tales datos;

Cabe destacar que los textos oficiales del Reglamento de Datos (2023) y el Reglamento de Gobernanza de Datos (2022), en una primera instancia, la traducción oficial al español se denominaba el *data holder* como titular de datos, no obstante, posteriormente se publicaron correcciones gramaticales de ambos reglamentos, donde se mencionaba que el término «titular de datos», en sus diversas formas gramaticales, se sustituye en todo el Reglamento por «tenedor de datos» en la forma gramatical adecuada y con las correspondientes adaptaciones gramaticales en el texto (Diario Oficial de la Unión Europea, 2024a, 2024b), lo cual abre la puerta al debate, sobre si únicamente tienen los datos y no son titulares de los mismos.

No obstante, continuando con este concepto, que ya se venía desarrollando en la doctrina con anterioridad, el WEF (2021c), indicó que, con la llegada de entidades de datos extremadamente grandes, como los proveedores de

plataformas, se generarían brechas de poder en la información y la negociación. Por esta razón, se requerían regulaciones para proteger la privacidad y garantizar una competencia justa. Plataformas como Apple y Google pueden no ser propietarias de los datos, pero en la práctica los controlan. El control excesivo del estado sobre el sector privado puede ocultar oportunidades para el uso democrático de datos y, como resultado, para la realización de valores en beneficio de la sociedad. Restringir a los poseedores de datos podría amenazar no solo el desarrollo económico, sino también llevar a la pérdida de su cooperación para lograr el interés público originalmente esperado. Dadas estas circunstancias, la gobernanza de datos debe dar forma de manera flexible al uso de los datos para incorporar a los poseedores de datos.

Cabe destacar que el Reglamento de datos (2023), establece que todo tratamiento de datos debe cumplir con el Derecho de la Unión en protección de datos, incluyendo la necesidad de una base jurídica válida según el RGPD. En específico, los titulares de datos deben poner los datos personales a disposición de los usuarios o terceros elegidos por el usuario, previa petición de este último. Se debe aplicar una política razonable de conservación de datos, en consonancia con el principio de limitación del plazo de conservación.

Además, se debe proporcionar información clara y transparente a los usuarios sobre el uso y tratamiento de datos, especialmente en contratos de prestación de servicios relacionados. Se debe hacer hincapié en que el reglamento de datos no otorga nuevos derechos a los titulares de datos para utilizar datos de productos o servicios relacionados; el uso de datos debe ser transparente y basarse en consentimiento informado, y por lo tanto, se garantiza el derecho de los usuarios a acceder y transferir sus datos, incluso en relaciones comerciales entre empresas.

En caso de corresponsabilidad del tratamiento de datos, los titulares de datos y los usuarios deben acordar sus respectivas responsabilidades de manera transparente y mutua. Los titulares de datos pueden aplicar medidas técnicas de protección para evitar la divulgación ilícita de datos, pero estas medidas no deben discriminar ni obstaculizar el acceso o la utilización de los datos por parte de los

usuarios. Aunado a lo anterior, los titulares de datos no deben explotar abusivamente su posición para obtener ventajas competitivas en mercados en los que puedan competir directamente con terceros.

De manera adicional, respecto a los titulares de los datos, el mismo Reglamento de Datos (2023) establece que los derechos y obligaciones de los titulares de datos, usuarios y terceros respecto al acceso, uso y disponibilidad de datos de productos y servicios relacionados. Los titulares de datos deben proporcionar a los usuarios acceso a los datos fácilmente disponibles de manera gratuita y segura. Los usuarios pueden compartir estos datos con terceros bajo ciertas condiciones, mientras que los terceros deben tratar los datos de manera acordada y respetar los derechos del usuario. Se establecen disposiciones para resolver disputas y se detallan criterios para la compensación por la puesta a disposición de datos entre las partes involucradas, con el objetivo de promover el intercambio justo y transparente de información.

En resumen, el Reglamento de Datos (2023) y Reglamento de Gobernanza de Datos (2022) introducen el concepto de tenedor de datos, definido como una persona física o jurídica con el derecho u obligación, según lo establecido en la normativa, de utilizar y poner a disposición datos, incluidos los datos del producto o servicios relacionados. Este nuevo rol surge en el contexto de una economía de datos en constante evolución, donde entidades extremadamente grandes tienen un control significativo sobre la información. Se destaca la necesidad de regulaciones para proteger la privacidad y garantizar una competencia justa. El reglamento establece que el tratamiento de datos debe cumplir con el Derecho de la Unión en protección de datos, incluyendo la transparencia en el uso y tratamiento de datos, el consentimiento informado y el derecho de los usuarios a acceder y transferir sus datos. Además, se detallan las responsabilidades y derechos de los titulares de datos, usuarios y terceros en relación con el acceso, uso y disponibilidad de datos de productos y servicios relacionados, con disposiciones para resolver disputas y promover el intercambio justo y transparente de información.

El tenedor de datos en una empresa tecnológica que utiliza datos generados por dispositivos IoT debe garantizar que los usuarios puedan acceder a sus datos de manera estructurada y transferible. Por ejemplo, si un cliente solicita datos de consumo energético de un dispositivo inteligente, el titular debe proporcionar esta información de forma segura y gratuita. Además, debe proteger los datos frente a accesos no autorizados e informar claramente a los usuarios sobre el tratamiento y uso de los datos compartidos con terceros, asegurando transparencia y cumplimiento normativo.

3.10 Usuario de datos (data user)

De acuerdo con el artículo 2 9) del Reglamento de Gobernanza de Datos (2022), un usuario de datos o *data user* es:

> "toda persona física o jurídica que tenga acceso legítimo a determinados datos personales o no personales y el derecho, incluido el que le otorga el Reglamento (UE) 2016/679 en el caso de los datos personales, a usarlos con fines comerciales o no comerciales"

Por otra parte, el Reglamento de Datos (2023), define el usuario o *user* como "una persona física o jurídica que posee un producto conectado o a la que se le han transferido por contrato derechos temporales de uso de dicho producto conectado, o que recibe servicios relacionados".

Las diferencias entre las definiciones de usuario de datos en el Reglamento de Gobernanza de Datos (2022) y el Reglamento de Datos (2023) radican en la naturaleza de los datos y los derechos asociados. Mientras que el primero abarca cualquier persona o entidad con acceso legítimo a datos personales o no personales, otorgándoles el derecho de utilizarlos con fines comerciales o no comerciales, el segundo se enfoca en aquellos que poseen o tienen derechos temporales de uso de productos conectados, sin especificar explícitamente el uso de datos.

Dicho lo anterior, para fines del presente apartado, se contempla el usuario de datos como la figura, porque tiene un vínculo hacia los datos y no un usuario que tiene únicamente un nexo con un producto o servicio.

Retomando el punto principal, el Reglamento de Gobernanza de Datos (2022), dentro de sus disposiciones establece que, los usuarios de datos son personas físicas o jurídicas que tienen acceso legítimo a datos personales o no personales y el derecho a utilizarlos con fines comerciales o no comerciales. Es fundamental asegurar un entorno competitivo para el intercambio de datos, donde los proveedores de servicios de intermediación de datos actúen como meros intermediarios y no utilicen los datos intercambiados para otros propósitos. Las condiciones para la prestación de servicios de intermediación de datos incluyen la neutralidad de los proveedores, la separación estructural de sus actividades y la utilización de datos facilitados únicamente para mejorar sus servicios. Además, se establecen requisitos específicos para proteger los derechos e intereses de los interesados y titulares de datos, como la transparencia en el tratamiento de datos personales y la notificación clara sobre el uso de datos en terceros países.

En conclusión, tanto el Reglamento de Gobernanza de Datos (2022) como el Reglamento de Datos (2023) definen al usuario de datos, aunque con enfoques ligeramente diferentes. Mientras que el primero abarca cualquier persona o entidad con acceso legítimo a datos personales o no personales y el derecho de utilizarlos con diversos fines, el segundo se centra específicamente en aquellos que poseen o tienen derechos temporales de uso de productos conectados. En el contexto de la gobernanza de datos, es esencial garantizar un entorno competitivo para el intercambio de datos, donde los proveedores de servicios de intermediación de datos actúen de manera neutral y transparente, respetando los derechos e intereses de los interesados y titulares de datos. Esto incluye condiciones para la prestación de servicios de intermediación de datos, como la neutralidad de los proveedores, la transparencia en el tratamiento de datos personales y la notificación clara sobre el uso de datos en terceros países.

Por ejemplo, un usuario de datos, como una empresa de análisis financiero que accede legítimamente a bases macroeconómicas mediante un proveedor de

datos, utiliza esta información para elaborar informes de mercado y asesorar a clientes. Para cumplir su rol, debe respetar los acuerdos de acceso, limitando el uso de los datos a los fines pactados y evitando compartirlos sin autorización. Además, debe garantizar que los datos sean precisos y actualizados, promoviendo decisiones informadas y fortaleciendo la confianza de sus clientes.

3.11 Productores de datos (Data producers)

De acuerdo con el Reglamento de Datos (2023), un consumidor o *consumer*, es "toda persona física que actúe con fines ajenos a su propia actividad comercial, negocio, oficio o profesión". No obstante, el término debe transmutar a las realidades de los tiempos contemporáneos, toda vez que existe un cambio del rol de consumidor hacia el de *prosumidor*.

Los actores y roles en los datos son fundamentales para identificar las formas de acceso y sus interacciones. Por ejemplo, la corriente académica presupone que la portabilidad de los datos es una medida en la encrucijada entre las leyes antimonopolio y de protección de datos, transmutando de consumidores a *prosumidores*, describiendo la urgencia de un instrumento jurídico que permita a los ciudadanos hacer un uso más eficaz de las oportunidades que ofrece la economía basada en datos. El usuario moderno, es tanto un consumidor como productor de contenido (datos en todas sus formas), por lo que se sugiere una intervención más específica para evitar que las redes sociales levanten una valla alrededor de su información común. Sin embargo, esto sólo puede garantizarse si la portabilidad está respaldada por la interoperabilidad con estándares abiertos, permitiendo a los *prosumidores* lograr una salida significativa, controlar su propio consumo o continuar con la captura de datos personales (Ursic, 2018, 76).

En este sentido, como ya se mencionó anteriormente cuando se abordaron los datos adquiridos en el apartado de tipos de datos; en este nuevo paradigma donde las personas ya no son únicamente consumidores, sino también *prosumidores*, el flujo de datos se convierte en un elemento vital y dinámico. Los

datos adquiridos no solo son el resultado de la interacción con productos y servicios, sino que también incluyen información generada, obtenida, comprada o licenciada de terceros a través de una variedad de medios, desde contratos comerciales hasta iniciativas gubernamentales abiertas.

Es crucial reconocer que estas interacciones conllevan obligaciones contractuales y legales que pueden influir en la reutilización y el intercambio de datos. En muchas ocasiones, los usuarios no podrían ser plenamente conscientes de que las aplicaciones que utilizan están vendiendo regularmente su información de ubicación a terceros, quienes a su vez la emplean para fines de *marketing* y otros propósitos. Estas empresas de terceros, a su vez, pueden intercambiar esos datos con agencias gubernamentales, creando un mecanismo de flujo de información.

Para abordar estas preocupaciones, diversas regulaciones como la Ley de Privacidad del Consumidor de California (CCPA por sus siglas en inglés) establecen medidas específicas. Por ejemplo, la CCPA requiere que las empresas que venden información personal proporcionen un enlace claro y vistoso en sus sitios web que permita a los usuarios enviar una solicitud de exclusión. Además, estas empresas no pueden obligar a los usuarios a crear una cuenta para ejercer su derecho a solicitar la exclusión de la venta de su información personal.

Este enfoque regulatorio busca empoderar a los *prosumidores*, brindándoles un mayor control sobre sus datos personales y fomentando la transparencia en las prácticas de recopilación y uso de datos por parte de las empresas. Sin embargo, el desafío radica en garantizar que estas regulaciones sean efectivas en un entorno digital en constante evolución, donde la privacidad y la protección de datos son aspectos fundamentales para el bienestar y la autonomía de los individuos.

Además, se debe subrayar que los datos se están categorizando como patrimonio cultural. Y si bien a *prima facie* no existe una titularidad de datos generados por máquinas, toda vez que las leyes actuales contienen un prominente antropocentrismo, con el tiempo esto puede cambiar y sistemas como IA podrían ser catalogados como productores de datos o algún tipo de figura jurídica similar.

Continuando la idea anterior, cada día los productores de datos desafían las nociones humanas de patrimonio al incluir elementos no humanos y más allá de lo humano. La transformación digital y el avance de la IA y los robots plantean cuestiones éticas sobre la equidad y los derechos de los robots como productores de datos; desde entidades biológicas hasta sistemas automatizados, una variedad de agentes contribuye al patrimonio cultural digital, ampliando así la noción de patrimonio más allá de lo humano hacia una convergencia *posthumana* (Cameron, 2021,230-272).

En resumen, el término "consumidor" definido por el Reglamento de Datos (2023) debe adaptarse a la realidad actual, donde los individuos no solo consumen datos, sino que también los producen como *prosumidores*. Esta transformación refleja un cambio en el flujo de datos, donde las personas son tanto consumidoras como productoras de contenido. Es fundamental garantizar que los *prosumidores* tengan control sobre sus datos y puedan ejercer sus derechos, especialmente en un entorno digital donde la privacidad y la protección de datos son cruciales. Las regulaciones como la Ley de Privacidad del Consumidor de California (CCPA) buscan empoderar a los *prosumidores* y promover la transparencia en las prácticas de recopilación y uso de datos. Además, el concepto de datos como patrimonio cultural está evolucionando para incluir elementos no humanos, como sistemas automatizados e IA, lo que plantea nuevas cuestiones éticas sobre la equidad y los derechos de los robots como productores de datos. Esta convergencia *posthumana* amplía la noción de patrimonio más allá de lo humano y desafía las nociones tradicionales de propiedad y responsabilidad.

Un productor de datos puede ser un influencer de redes sociales que genera contenido multimedia a diario, como videos, publicaciones y transmisiones en vivo, que recopilan datos como métricas de interacción, comentarios y preferencias de los usuarios. Este experto utiliza herramientas de análisis de redes sociales para optimizar su contenido en función de los datos generados, como identificar los horarios de mayor actividad o los temas más comentados. Para desempeñar su rol de manera ética, el influencer implementa prácticas transparentes al informar a sus seguidores sobre el uso de sus datos, respeta las

configuraciones de privacidad de las plataformas y asegura que cualquier colaboración con marcas incluya cláusulas que protejan la privacidad de los datos compartidos.

3.12 Propietario de los datos (Data owner)

De manera puntual cuando se alude al concepto de propietario de los datos, este tiene un vínculo con la propiedad de los datos *per se*. No obstante, antes de comenzar con este apartado, se debe hacer la aclaración que, aunque a *prima facie*, pudiera inferirse que un propietario de los datos o *data owner* es el interesado o *data subject* de datos personales (Tang, 2023, 441), en realidad, la propiedad de los datos puede recaer en diversas entidades o individuos dependiendo del contexto y las disposiciones legales.

Otro aclaración necesaria, es subrayar que los conceptos de *data holder* y *data owner* pudieran ser considerados términos equivalentes por algunos autores (Torra, 2017, 56), pero a *contrario sensu*, otros autores argumentan que cada persona que posee datos es el tenedor de los datos (*holder*), pero el tenedor de los datos no es necesariamente el propietario de los datos (*owner*); es decir, para un solo dato, existen cuatro titulares diferentes: titular del dato, que posee todos los derechos y dominio sobre el dato; productor de datos, quien está autorizado por el propietario para crear y recopilar los datos; usuario de los datos, que está autorizado por el titular para utilizar los datos; Agente de datos, quien está autorizado por el titular para poseer una especie de derechos sobre los datos (Miao et al., 2022, 124).

Sin perjuicio del debate entre *data holde*r y *data owner*, de manera puntual un *data holder* es responsable de mantener y gestionar los datos, y un *data owner* es la entidad que posee los derechos de propiedad sobre los datos y puede retener ciertas responsabilidades legales incluso después de compartir los datos con terceros. Cabe destacar que también los términos ***Data Governor***

(Gobernador de datos) y ***Data Owner*** (Propietario de datos) se utilizan indistintamente en muchas organizaciones como equivalentes (Plotkin, 2021, 7).

Independientemente de lo anterior y retomando el asunto principal, los propietarios de datos tienen responsabilidades relativas a los datos que caen dentro de su ámbito y, por lo tanto, deben preocuparse por los datos más importantes para ellos porque, francamente, nadie más lo hará. Por ejemplo, el jefe de asuntos legales y los sistemas que utiliza almacenan datos legales, esos datos son su preocupación, empero, no significa que le pertenezca (aunque eso podría ser semántico), pues significa que se preocupa por su calidad, integridad, disponibilidad y más, incluso teniendo la autoridad sobre quién puede acceder a los datos (Reichental, 2023, 125).

El Propietario de Datos es un ejecutivo *senior* un nivel por debajo del *C-level*, responsable de los datos generados en su área. Su rol es estratégico, estableciendo metas y directrices para el manejo de datos. Supervisa al personal operativo y es la última instancia en la resolución de discrepancias en el manejo de datos. Garantiza la capacidad de la organización para tomar decisiones sobre datos (Bollweg, 2022, 26).

En síntesis, una definición breve de propietario de datos es la persona que es responsable de un activo de datos, debido a los numerosos procesos que pueden afectar la calidad de los datos, el propietario de los datos se encontrará en una posición relativamente impotente. En general, un mejor enfoque es garantizar que se identifique a los propietarios de los procesos y que ayuden a garantizar la calidad de los datos creados y actualizados por los procesos de su propiedad. Esto significará que, si se decide asignar propietarios de datos, estos tendrán más control sobre los datos de los que son responsables al trabajar en asociación con el propietario del proceso correspondiente (King & Schwarzenbach, 2020, 101).

En conclusión, el concepto de propietario de datos es esencial en la gestión de la información, aunque su definición puede variar según el contexto y las disposiciones legales. Si bien hay un debate sobre la equivalencia entre "*data holder*" y "*data owner*", ambos desempeñan roles cruciales en la protección y calidad de los datos. Los propietarios de datos, mayormente ejecutivos senior,

tienen la responsabilidad de establecer directrices y metas para el manejo de datos en sus áreas respectivas, trabajando en colaboración con propietarios de procesos para garantizar la integridad y utilidad de la información. Asignar propietarios de datos puede mejorar el control y la calidad de los datos en una organización, asegurando así la toma de decisiones informadas y la protección de los activos de datos.

Un propietario de datos, como un director de Recursos Humanos, gestiona los datos personales de los empleados, asegurando su precisión, actualización y acceso controlado. Por ejemplo, puede establecer políticas que limiten el acceso a datos salariales al equipo financiero y personal autorizado de RR. HH., cumpliendo regulaciones como el RGPD. También colabora con el equipo de TI para implementar medidas de seguridad, como cifrado y monitoreo, protegiendo la información frente a accesos no autorizados.

3.13 Fideicomisario de datos (Data trustee)

El Fideicomisario de Datos es en última instancia responsable y el guardián principal de un dominio de datos específico, definiendo el alcance de los datos dentro del mismo, pero típicamente delegaría la gestión de los datos a los administradores de datos para su manejo dentro de dicho dominio (Datoo, 2019, 78-80).

Un fideicomisario o fiduciario de datos es un individuo o entidad responsable de garantizar la precisión, actualidad y calidad de cada elemento de datos dentro de una organización o empresa. A diferencia de un administrador, cuyas responsabilidades también pueden incluir estandarización y definición, el enfoque principal de un fideicomisario de datos es mantener la integridad de los datos eliminando esfuerzos y recursos redundantes y asegurando la precisión y actualización de los datos. El fideicomisario de datos es responsable de cumplir con los requisitos de calidad, brindar confianza al usuario en los datos e identificar la verdadera fuente de los datos. Este rol es crucial para resolver problemas de

propiedad de datos y facilitar el acceso a los datos para todos los usuarios dentro de la organización (Harjika, 2022, 67-68).

A medida que crece el grado de intercambio de datos y las unidades de negocio dependen de información común, se vuelve esencial que sólo el administrador de los datos tome decisiones sobre el contenido de los datos y será responsable de cumplir con los requisitos de calidad impuestos a los datos de los cuales es responsable; como resultado de compartir datos en toda la empresa, el fideicomisario es responsable de la exactitud y vigencia de los elementos de datos designados y, posteriormente, debe reconocer la importancia de esta responsabilidad de tutela (Ghavami, 2020, 50).

Varias empresas proveedoras de servicios en la nube han comenzado a abrir nuevos centros de datos en Alemania debido a la demanda de clientes alemanes de que sus datos sean alojados exclusivamente en dicho país. Microsoft es la primera empresa de servicios en la nube pública en implementar un modelo de acceso a datos en el que un Fideicomisario de Datos alemán controla el acceso a los datos del cliente de acuerdo con la ley alemana (Bendiek, 2017, 696).

Como ejemplo, en algunos contratos se puede encontrar la cláusula de que un *Data trustee*, significa un organismo que recopila o procesa datos, verbigracia de vehículos (por ejemplo, información de software), de diferentes fabricantes y proveedores y los pone a disposición de terceros autorizados, por ejemplo, para pruebas de aptitud para la circulación, monitoreo del consumo de combustible a bordo, monitoreo de campo y otros casos de uso (UNECE, 2022, 2).En pocas palabras, se debe determinar que el *data trustee* es sinónimo de tutelar datos, es decir una transición de propiedad hacia tutelar.

En resumen, el papel del Fideicomisario de Datos es fundamental en la gestión y protección de la información en una organización, ya que es el principal responsable de garantizar la precisión, integridad y calidad de los datos dentro de un dominio específico. Aunque típicamente delega la gestión de los datos a los administradores de datos, su enfoque principal radica en mantener la integridad de los datos, eliminando redundancias y asegurando la precisión y actualización de la información. Esta figura se vuelve aún más crucial en entornos donde el

intercambio de datos es frecuente, y su rol es clave para resolver problemas de propiedad de datos y facilitar el acceso a la información para todos los usuarios dentro de la organización. Además, ejemplos como el modelo implementado por Microsoft en Alemania demuestran cómo el concepto de Fideicomisario de Datos está siendo reconocido y adoptado en diferentes contextos para garantizar la protección y cumplimiento de las leyes locales en cuanto al manejo de datos.

Un fideicomisario de datos en una empresa de transporte supervisa los datos generados por una flota de vehículos conectados, como mantenimiento, consumo de combustible y rutas. Es responsable de garantizar su precisión, actualización y cumplimiento normativo. Por ejemplo, colabora con fabricantes y proveedores para centralizar la información en una plataforma segura, permitiendo el acceso únicamente a terceros autorizados, como organismos de certificación, para evaluar el rendimiento ambiental. Además, implementa procesos para eliminar duplicidades y mantener la calidad y confiabilidad de los datos para usuarios internos y externos.

3.14 Administradores de datos (Data Stewards)

Por su parte, los administradores de datos tienen la responsabilidad de gestionar los derechos y la reutilización de los datos de manera responsable. Actúan como promotores de la confianza tanto dentro como fuera de la organización, buscando oportunidades de colaboración entre sectores y respondiendo proactivamente a solicitudes externas de acceso a datos. Su papel incluye establecer y demostrar su confiabilidad a través de roles profesionales y certificaciones éticas. Además, pueden facilitar la relación entre el sector público y privado en el intercambio de datos, contribuyendo así a una gestión más responsable y a relaciones sectoriales más sólidas (WEF, 2022, 14).

El rol de *data steward* es fundamental en la gobernanza de datos. Tienen la responsabilidad de gestionar los datos diariamente en la organización, siendo expertos en los conjuntos de datos que supervisan. Trabajan con otros *stewards*

para tomar decisiones sobre el uso de datos, mejorar la calidad estos y solucionar problemas relacionados con los datos. Aunque no son personal técnico, tienen habilidades especializadas en datos y conocen profundamente los datos dentro de su alcance. Pueden formar parte del Data Governance Board (DGB) y también pueden crear su propio Data Steward Council (DSC). Sus responsabilidades incluyen identificar y definir activos de datos, colaborar en la creación de procesos y políticas, garantizar la calidad y seguridad de los datos, resolver problemas en el ciclo de vida de los datos y representar los intereses de los propietarios de datos en el DGB (Reichental, 2023, 126-127).

Tomando en cuenta algunos principios inspirados en el marco de referencia TOGAF, un *Data trustee* es responsable de la precisión y actualidad de los datos, mientras que las responsabilidades de un *Data Stewards* son más amplias e incluyen la estandarización de datos, políticas de uso de datos, gestión de acceso a datos y tareas de definición. El *Data Steward* es responsable de garantizar el uso adecuado de los datos de acuerdo con las reglas de acceso, las regulaciones y cumplimiento, la soberanía de los datos, las reglas jurisdiccionales de los datos y los acuerdos contractuales con clientes y/o proveedores de datos externos. Esto implica que puede ser necesario un cambio cultural de la ***ownership* (propiedad)** de los datos a la ***trusteeship* (tutela)** y la ***stewardship* (administración)** de los datos (Ghavami, 2020, 50).

Cabe hacer la aclaración que, en ocasiones el ***Data Stewards*** es considerado un símil al término ***Data Manager***, pero también por algunos autores consideran que la diferencia clave entre un *data manager* y un *data steward* radica en sus enfoques y responsabilidades dentro de una organización en relación con la gestión de datos. Mientras que el *data manager* se centra en la gestión práctica de los datos, asegurando su calidad y disponibilidad para el uso organizacional, el *data steward* tiene como objetivo promover y proteger el uso adecuado de los datos, actuando como un defensor y asesor en cuestiones relacionadas con su uso ético y autorizado dentro de la organización, interactuando más con las personas que con los datos en sí (Holt, 2021, 122).

Incluso, otros autores consideran que los *data managers* designan a los *data stewards*. Es decir, que el *Data Manager* es responsable de supervisar los activos y procesos de datos dentro de un dominio, alineando los esfuerzos comerciales y de TI, tomando decisiones sobre los datos y designando *Data Stewards* y Grupos de Trabajo *Ad-Hoc* para resolver problemas y recomendar políticas. Representa su dominio en el Equipo de Gestión de Datos para facilitar la comunicación y colaboración entre dominios. Por otro lado, el *Data Steward* utiliza su conocimiento empresarial y técnico para desarrollar reglas y definiciones comerciales, identificar problemas de cumplimiento y hacer recomendaciones de políticas, bajo la dirección del *Data Manager* y la orientación de la Gestión de Gobierno de Datos, siendo reconocido como la persona de referencia para problemas y conocimientos de datos dentro de su dominio (Hopper, 2023, 99).

No obstante, si bien los autores anteriores han distinguido entre data manager y *data steward*, de acuerdo a lo escrito hasta el momento, este rol de control/mando se superpone con la armonía de roles redactada en este capítulo relacionada con *data owners* o *data trusteed* designando *data stewards*. Por lo tanto, para obtener una armonía dentro de una arquitectura empresarial con un diseño, planificación, implementación y gobierno, se tomará TOGAF como guía oficial y considerar el *data steward* como *data manager*, sin menoscabo de que en alguna inmensa organización sea necesario un rol de coordinación hacia varios *stewards.* En síntesis*,* se debe determinar que el administrador de datos significa administrar los datos, es decir siguiendo las ideas anteriores, es la transición de propiedad a tutelar y en el caso, hacia administrar.

En conclusión, el papel de los administradores de datos, *Data Stewards*, es esencial en la gestión eficaz de la información dentro de una organización. Son responsables de garantizar la precisión, integridad y calidad de los datos, promoviendo la confianza tanto dentro como fuera de la organización y facilitando la colaboración entre sectores. Los *Data Stewards* tienen un profundo conocimiento de los conjuntos de datos que supervisan y trabajan en estrecha colaboración con otros para tomar decisiones sobre el uso de los datos, mejorar su calidad y resolver problemas relacionados con ellos. Aunque hay cierta

superposición de roles y definiciones, es crucial reconocer el papel fundamental que desempeñan los administradores de datos en la gestión responsable de la información en una organización, facilitando así el cumplimiento de las regulaciones, la protección de los activos de datos y la promoción de relaciones sectoriales sólidas.

Por ejemplo, un administrador de Datos en una institución bancaria supervisa los datos de clientes, como cuentas y transacciones, estableciendo políticas claras para el acceso adecuado a información sensible. Colabora con equipos de TI para proteger los datos frente a brechas de seguridad y con los propietarios de datos para definir estándares de calidad, como mantener actualizados los datos de contacto. Ante inconsistencias, como duplicados, toma medidas correctivas para garantizar la precisión y confiabilidad de los datos, cumpliendo regulaciones como el RGPD y fortaleciendo la confianza de los clientes.

3.15 Custodio de datos (Data custodians)

Un *data custodian* es una entidad encargada de manejar datos de forma segura, especialmente en situaciones donde se requiere proteger la identidad de los datos. El *data custodian* actúa como un intermediario confiable para garantizar el manejo seguro y controlado de los datos, facilitando el acceso autorizado mientras protege la privacidad y confidencialidad de la información (ENISA, 2021, 32-34).

Cabe hacer la aclaración que, el **propietario de los datos** suele ser un líder senior dentro de una función empresarial con la responsabilidad final de la estrategia, la calidad, el valor y las aprobaciones de acceso a los datos de su dominio (Reichental, 2023, 127), mientras que el **fideicomisario de datos** es responsable último y principal "guardián" de un dominio de datos específico, definiendo su alcance, supervisando su estado y estableciendo la hoja de ruta estratégica para el dominio, y aunque son responsables directos de los datos, típicamente delegan las responsabilidades de gobernanza diaria a los

administradores y custodios de datos. Al respecto, el **administrador de datos** es un experto en la materia que define tipos de categorías de datos, valores permitidos y requisitos de calidad de datos, mientras que los **custodios de datos** son responsables desde una perspectiva tecnológica, gestionando los derechos de acceso y aplicando controles para garantizar la integridad, seguridad y privacidad de los datos (The Law Society, 2023, 131-132). En síntesis, siguiendo las ideas anteriores, se debe determinar que el custodiador de datos significa cuidar los datos, es decir, es una transición de propiedad a tutelar y de administrar hacia custodiar.

En resumen, el papel de los *data custodian* es vital en la gestión segura de los datos, especialmente en entornos donde se requiere proteger la identidad de la información. Actuando como un intermediario confiable, el *data custodian* garantiza el manejo seguro y controlado de los datos, facilitando el acceso autorizado mientras salvaguarda la privacidad y confidencialidad de la información. Aunque el propietario de los datos y el fideicomisario de datos tienen responsabilidades finales sobre la estrategia y el cuidado general de los datos, respectivamente, suelen delegar las tareas de gobernanza diaria a los administradores y custodios de datos. Mientras que los administradores definen categorías de datos y requisitos de calidad, los custodios de datos se centran en la aplicación de medidas tecnológicas para garantizar la integridad, seguridad y privacidad de los datos.

Por ejemplo, un Custodio de Datos en una universidad gestiona la seguridad y privacidad de información estudiantil, como calificaciones y expedientes. Configura permisos en sistemas de gestión académica para garantizar el acceso exclusivo a personal autorizado, como profesores y servicios escolares. Además, implementa cifrado para proteger los datos almacenados y diseña procedimientos de recuperación ante desastres para asegurar su accesibilidad en casos de fallos técnicos o ciberincidentes. Actuando como enlace técnico, cumple con las políticas del Data Owner y asegura el cumplimiento normativo en la protección de los datos estudiantiles.

A continuación, se desarrolla una tabla que pueden definir las principales tareas de los *data owner, data trustee, data steward* y *data custodian.*

Cuadro 2. Principales tareas entre *data owner, data trustee, data steward y data custodian.*

Data Owner.	**Data trustee**	**Data Steward:**	**Data Custodian:**
• Calidad. • Responsable. • Supervisión. • Autorización de acceso. • Estrategia. • Aplicación.	• Legalidad. • Responsabilidad fiduciaria. • Gestión de riesgos. • Cumplimiento normativo. • Confidencialidad.	• Calidad. • Creación de procesos y estándares. • Cumplimiento. • Protección. • Resolver problemas.	• Calidad. • Almacenamiento. • Seguridad. • Controles de acceso. • Arquitectura. • Bases de datos. • Recuperación de desastres.

Cuadro basado en Reichental (2023, 127).

3.16 Arquitecto de Datos (Data Architect)

Un arquitecto de datos es un profesional encargado de diseñar, crear, implementar y gestionar la arquitectura de datos de una organización. La arquitectura de datos consiste en un conjunto de reglas, políticas, estándares y modelos que rigen y definen el tipo de datos recopilados y cómo se utilizan, almacenan, gestionan e integran dentro de la organización y sus sistemas de datos. El arquitecto de datos es fundamental para garantizar la eficiencia y la alta productividad en los equipos de ciencia de datos, ya que define cómo se almacenarán, consumirán, protegerán, integrarán y gestionarán los datos por parte de diferentes entidades de datos y sistemas de TI, así como cualquier aplicación que utilice o procese esos datos de alguna manera. Aunque no suele ser un miembro permanente de un solo equipo de ciencia de datos, el arquitecto de datos trabaja estrechamente con varios

equipos para asegurar que se cumplan los objetivos de la organización en términos de gestión y aprovechamiento de datos (Jägare, 2019, 159).

Como analogía, los arquitectos de datos organizan y estructuran los datos de una empresa de manera similar a cómo una biblioteca indexa sus libros. Su función es crear modelos conceptuales que ayuden a entender quién es responsable de qué información. Esto incluye la gestión de datos maestros y modelos de datos (Carruthers & Jackson, 2021, 97). Este puesto desarrolla diseños lógicos y físicos de modelos de datos para aplicaciones. Con suficiente experiencia, esta persona también puede diseñar la arquitectura de datos general de una organización (Gregory, 2018, 34).

Además, los arquitectos de datos suelen trabajar con los ingenieros de datos para crear soluciones que optimicen el rendimiento de las bases de datos, las aplicaciones analíticas y los marcos (Sozonov, 2020, 19).

Los arquitectos de datos deben reunirse con los equipos de gestión y ayudar a determinar el método para recopilar los datos necesarios y cómo organizarlos. El arquitecto de datos revisa y analiza el entorno de datos de una organización, se ocupa de la planificación de futuras bases de datos y ayuda a investigar e implementar soluciones técnicas para gestionar los datos para su organización y sus usuarios de datos (Klidas & Hanegan, 2022, 102). Sin embargo, la arquitectura de datos requiere paciencia, atención a los detalles y la capacidad de escuchar y combinar múltiples puntos de vista en un modelo de sistema coherente. Un buen arquitecto de datos también debe poder alejarse del argot técnico y simplificar el lenguaje para que cualquiera pueda entender cómo funciona (Woodard, 2020, 90).

En resumen, el arquitecto de datos desempeña un papel importante en el diseño, implementación y gestión de la arquitectura de datos de una organización. Esta arquitectura establece reglas y estándares para el uso, almacenamiento, gestión e integración de datos dentro de la organización y sus sistemas. Similar a un bibliotecario que indexa libros, el arquitecto de datos organiza la información para comprender quién es responsable de qué datos, desarrollando modelos conceptuales y físicos para su gestión. Trabajando estrechamente con diversos

equipos, los arquitectos de datos optimizan el rendimiento de bases de datos y aplicaciones analíticas. Además, colaboran con el equipo de gestión para determinar los métodos de recopilación y organización de datos necesarios. Y, por lo tanto, la habilidad para simplificar el lenguaje técnico y escuchar múltiples puntos de vista es fundamental para un arquitecto de datos exitoso, cuya labor es fundamental para garantizar la eficiencia y productividad en el manejo de datos dentro de la organización.

Por ejemplo, un Arquitecto de Datos en un hospital diseña un sistema centralizado para gestionar expedientes médicos electrónicos. Crea un modelo lógico que integra datos de diversas fuentes, como historiales clínicos y resultados de laboratorio, asegurando su correcta indexación y disponibilidad para personal autorizado. También diseña la arquitectura física para optimizar el rendimiento, garantizando acceso rápido en emergencias y colabora con ingenieros de datos para implementar una base de datos segura y escalable y con el equipo de seguridad para cumplir normativas como HIPAA, mediante protocolos de cifrado y acceso controlado.

3.17 Ingeniero de Datos (Data Engineer)

La función principal del ingeniero de datos es garantizar la disponibilidad y consistencia de los datos para los científicos de datos. Esto implica la creación de pipelines de datos (canalización de datos) consistentes y fácilmente accesibles para el consumo por parte de los científicos de datos. Los ingenieros de datos se encargan de la ingesta, procesamiento y almacenamiento de datos, asegurando que estos procesos sean invisibles para los científicos de datos. Aunque no necesitan conocimientos en aprendizaje automático o estadísticas, deben tener habilidades técnicas para manejar aspectos complejos de la ingeniería de datos (Jägare, 2019, 157-158).

Incluso se podría hacer la analogía de que un científico de datos es un piloto de carreras, mientras que el ingeniero de datos es el constructor de los

autos de carrera; por lo tanto, aunque son menos visibles, los ingenieros de datos son esenciales para asegurar la calidad y disponibilidad de los datos, además la ingeniería de datos es un campo amplio, pero un ingeniero de datos no necesita dominar todos los aspectos, sino enfocarse en transformar los datos para análisis (Smith, 2020, 29-30).

La ingeniería de datos se centra en construir y mantener sistemas para superar problemas de procesamiento y manejo de grandes volúmenes de datos. Los ingenieros de datos utilizan habilidades en informática y *software* para diseñar sistemas que manejen eficientemente grandes conjuntos de datos. Su objetivo es crear plataformas de procesamiento de datos escalables y modulares, facilitando así el trabajo de los científicos de datos para obtener ideas y análisis útiles (Pierson, 2021, 27-28).

Cabe destacar que algunos autores manejan una diferencia entre ingeniero de datos e ingeniero de sistemas de datos. Al respecto, el ingeniero de datos adquiere, conecta, prepara y gestiona datos, especialmente para sistemas de gestión y preparación de datos a gran escala, multi-fuente y en tiempo real. Mientras tanto, el ingeniero de sistemas de datos programa y desarrolla prototipos y plataformas de sistemas, integrando modelos de datos y lógicas analíticas en el espacio de información empresarial y sistemas de apoyo a la toma de decisiones. La principal diferencia radica en que el ingeniero de datos se enfoca en la gestión y preparación de datos, mientras que el ingeniero de sistemas de datos se concentra en el desarrollo de sistemas que incorporen modelos de datos y lógicas analíticas para el soporte a la toma de decisiones (Cao, 2018, 327).

En conclusión, el ingeniero de datos desempeña un papel fundamental en la garantía de la disponibilidad y consistencia de los datos para los científicos de datos, siendo esenciales en la creación de *pipelines* de datos consistentes y fácilmente accesibles. Aunque menos visibles que los científicos de datos, los ingenieros de datos son comparables a los constructores de autos de carrera, asegurando la calidad y disponibilidad de los datos para su análisis. Su labor se centra en construir y mantener sistemas eficientes que manejen grandes volúmenes de datos, facilitando así el trabajo de los científicos de datos para

obtener ideas y análisis útiles. Aunque existe una distinción entre ingenieros de datos e ingenieros de sistemas de datos, ambos desempeñan un papel crítico en la gestión y preparación de datos, así como en el desarrollo de sistemas que apoyen la toma de decisiones empresariales.

Por ejemplo, un Ingeniero de Datos (*Data Engineer*) en una empresa de comercio electrónico diseña un sistema de canalización de datos que extrae información de transacciones en tiempo real desde la plataforma de ventas, limpia los datos para eliminar duplicados y errores, y los almacena en un lago de datos (*data lake*) centralizado. Además, configura un sistema escalable en la nube para procesar millones de registros diariamente, asegurando que los datos estén listos y accesibles para los científicos de datos que analizarán patrones de compra y tendencias de clientes.

3.18 Científico de Datos (Data Scientist)

Con respecto a los científicos de datos, se debe definir primero que la ciencia de datos se centra en descubrir conocimiento a través del análisis de datos, empleando técnicas matemáticas y algoritmos para resolver problemas complejos en negocios y ciencia. Los científicos de datos aplican métodos predictivos para obtener información valiosa, optimizando procesos empresariales y logrando objetivos científicos específicos. Para ser un científico de datos se requiere experiencia en matemáticas, estadísticas, programación y en el área de interés, y específicamente, el aprendizaje automático es una herramienta clave en esta disciplina, permitiendo hacer predicciones automatizadas a partir de datos (Pierson, 2021, 25).

Un científico de datos desarrolla modelos matemáticos para predecir, requiere formación avanzada en informática, matemáticas o estadística, y habilidades de programación. Además de ser técnicamente competente, necesita pensamiento crítico y comprensión de los objetivos comerciales. A veces se diferencia del analista de datos, quien se centra en recopilar, interpretar y analizar

datos para extraer conclusiones. Ambos roles requieren dominio de lenguajes como R, Python, SQL y C (Jägare, 2019, 157).

El científico de datos realiza descubrimientos y exploraciones científicas sobre datos, resolviendo problemas impulsados por datos. Su rol se puede especificar como modelador de datos, minero de datos, analista de aprendizaje automático, analista estadístico, analista de optimización, analista de pronósticos, analista predictivo, analista de visualización, analista cuantitativo y analista de *insights* (Cao, 2018, 327).

Por otra parte, no se debe confundir a los científicos de datos con los profesionales de inteligencia empresarial (BI), ya que los segundos interactúan con los interesados en el negocio para entender sus necesidades de informes, mientras que los científicos de datos tienen un conocimiento más profundo de datos y análisis; tienden a trabajar más con *big data*, utilizan herramientas más sofisticadas y colaboran con una variedad más amplia de roles técnicos y comerciales en una organización que los profesionales de BI. Aunque ambos manipulan herramientas específicas, un buen científico de datos también comprende el dominio subyacente y el negocio, colaborando eficazmente con expertos en el campo para lograr un impacto más amplio en el descubrimiento de conocimiento accionable (Cao, 2018, 319-350).

Cabe destacar que, los datos son fundamentales en proyectos de ciencia de datos, pero su acceso no garantiza su uso legal o ético; los científicos de datos deben entender estas regulaciones y tener una comprensión ética para utilizar los datos de manera adecuada (Kelleher & Tierney, 2018, 21). Los marcos legales varían, pero la mayoría de los países comparten leyes antidiscriminatorias y de protección de datos personales, estableciendo principios como la limitación de la recolección de datos, la calidad de estos, la especificación del propósito, la seguridad de los datos y la transparencia (Kelleher & Tierney, 2018, 205-208). En la práctica, los científicos de datos podrían tomar datos proveídos u observados para convertirlos en datos derivados o inferidos, los cuales ya fueron definidos en el capítulo relativo a tipos de datos.

En resumen, los científicos de datos desempeñan un papel crucial en la extracción de conocimiento a partir del análisis de datos, empleando técnicas matemáticas y algoritmos para resolver problemas complejos en diversos campos. Requieren una amplia gama de habilidades en áreas como matemáticas, estadísticas, programación y dominio del área de interés, con el aprendizaje automático como una herramienta clave en su arsenal. A diferencia de los analistas de datos, los científicos de datos se centran en descubrir patrones y tendencias a través de la exploración científica de los datos, utilizando una variedad de roles como modeladores, analistas y mineros de datos. Es crucial distinguirlos de los profesionales de inteligencia empresarial, ya que los científicos de datos tienen un conocimiento más profundo de los datos y tienden a trabajar con *big data* y herramientas más sofisticadas. Además, deben tener en cuenta las regulaciones legales y éticas en el uso de los datos para garantizar su aplicación apropiada en proyectos de ciencia de datos.

Por ejemplo, un Científico de Datos (*Data Scientist*) en una empresa de transporte público, analiza datos históricos de los horarios y la ocupación de autobuses para predecir la demanda de pasajeros en diferentes rutas y horarios. Puede desarrollar un modelo de aprendizaje automático que identifica patrones en el comportamiento de los pasajeros según factores como el clima, eventos locales y días festivos. Luego, colabora con el equipo de operaciones para implementar un sistema que ajusta automáticamente la frecuencia de los autobuses en tiempo real, optimizando recursos y mejorando la experiencia del usuario. Además, garantiza que los datos utilizados respeten las normativas de privacidad y sean tratados de forma ética.

3.19 Analista de Datos (Data Analyst)

Por otra parte, como se mencionó anteriormente la analítica de datos es crucial para extraer valor de conjuntos de datos crudos, requiriendo herramientas y

técnicas específicas para generar conclusiones y respuestas (Reichental, 2023, 51-52).

Un *Data Analyst* es un profesional encargado de examinar extensas cantidades de datos con el fin de facilitar la toma de decisiones empresariales. Utilizan herramientas como SQL para identificar patrones relevantes en los datos y así responder interrogantes comerciales complejas. Este rol demanda habilidades tanto para comprender como para comunicar eficazmente la información derivada de los datos y su impacto en la empresa, ocupando una posición intermedia entre los aspectos teóricos y prácticos de la Ciencia de Datos (Chojecki, 2020, 30).

Por otro lado, desde un enfoque de privacidad, el analista de datos utiliza datos anonimizados para realizar actividades de minería de datos, como predicción y descubrimiento de conocimientos. Siguiendo algunas regulaciones gubernamentales, solo se puede utilizar datos anonimizados para la minería de datos, sin menoscabo del consentimiento o acuerdo de adhesión. Por lo tanto, es crucial que los datos proporcionados sean compatibles con las funcionalidades de minería de datos (Venkataramanan & Shriram, 2017, 5).

Cabe destacar que dentro de esta clasificación puede desprenderse el rol de **Analista de calidad de datos** (*Data quality analyst*), siendo el responsable de auditar, monitorear y medir la calidad de los datos diariamente, y recomendar acciones para corregir y prevenir errores y defectos (Mahanti, 2019, 410), así como **Analista de seguridad de datos** (Data Security Analyst), que protege la información en computadoras y redes (Adams, 2024, 41).

En resumen, la analítica de datos es esencial para desentrañar información valiosa de conjuntos de datos complejos, lo que implica el empleo de herramientas y técnicas especializadas para generar respuestas y conclusiones significativas. El rol del *Data Analyst* se destaca como fundamental en este proceso, ya que se encarga de explorar grandes volúmenes de datos para respaldar la toma de decisiones empresariales mediante la identificación de patrones relevantes. Además, el enfoque en la privacidad subraya la importancia de utilizar datos anonimizados y cumplir con regulaciones gubernamentales para llevar a cabo actividades de minería de datos de manera ética. Dentro de esta clasificación,

surgen roles específicos como el Analista de calidad de datos y el Analista de seguridad de datos, que contribuyen a garantizar la integridad y protección de la información.

Por ejemplo, un Analista de Datos (*Data Analyst*) en una empresa de comercio electrónico utiliza herramientas como *SQL* y *Excel* para analizar datos de ventas mensuales. Identifica que ciertos productos tienen un aumento en ventas durante días festivos específicos. Presenta estos hallazgos en un informe visual, destacando las oportunidades de promociones dirigidas en esas fechas. Además, recomienda ajustar el inventario para optimizar los ingresos, asegurándose de que todos los datos analizados cumplan con las regulaciones de privacidad al estar anonimizados.

Finalmente, a continuación, se hace una tabla que específica las principales funciones entre los roles abordados como *Data Architect*, *Data Engineer*, *Data Scientist* y *Data Analyst*:

Cuadro 3. Principales tareas entre *data architect, data engineer, data scientist y data analystn.*

Data Architect	**Data Engineer**	**Data Scientist**	**Data Analyst**
● Diseño e implementación de la arquitectura de datos. ● Establecimiento de reglas y estándares para la gestión de datos. ● Colaboración con equipos para cumplir	● Garantía de disponibilidad y consistencia de datos. ● Creación de pipelines de datos eficientes. ● Ingesta, procesamiento y almacenamiento de datos.	● Uso de técnicas matemáticas y algoritmos para análisis de datos. ● Aplicación de métodos predictivos para optimizar procesos empresariales. ● Experiencia	● Análisis de grandes conjuntos de datos para toma de decisiones. ● Identificación de patrones relevantes mediante herramientas como SQL. ● Posición intermedia

objetivos de gestión y aprovechamiento de datos.		en matemáticas, estadísticas y programación.	entre teoría y práctica de la Ciencia de Datos.

3.20 Oficiales/Directores vinculados a los datos (C-level)

El papel del Director de Datos o ***Chief Data Officer*** (CDO), ha ganado prominencia en los últimos años como la persona responsable de los resultados de los datos, aunque también son reconocidos como Directores de Datos Digitales (***Chief Digital Officers***) o Director Global de Datos (***Global Head of Data***), pero independientemente del título, se entiende claramente que el papel y el patrocinio son críticos para la gestión y gobernanza de los datos empresariales (Fryman et al., 2016, 33).

Además del CDO, han surgido nuevos roles como el Director de Análisis (**Chief Analytics Officer**), el Director de Gobernanza de Datos (**Chief Data Governance Officer**) y el Director de Riesgos de Datos (**Chief Data Risk Officer**), reflejando la creciente importancia de la gestión de datos en las organizaciones, pero cada uno de estos nuevos títulos es en realidad una rama de una de las responsabilidades principales de la generación original de CDO (Fryman et al., 2016, 34-35).

A la par, también surgen los oficiales de ética de datos (**Data ethics officers**) y sus respectivos comités de ética de IA, ya sea centralizando la gestión ética en un oficial específico o en un equipo; la gestión de ética de datos se distingue por su enfoque estratégico en lugar de centrarse únicamente en el cumplimiento de leyes de privacidad, buscando construir confianza y prepararse para regulaciones futuras (Hirsch et al., 2023, 62). O también surge el Oficial de Cumplimiento de Datos (**Data Compliance Officer**) (Sensen et al., 2024, 88), y

como lo dice su nombre, se encarga de garantizar que una organización cumpla con las regulaciones y normativas relacionadas con la gestión de datos.

Cabe destacar que, como ya se mencionó en otro apartado del presente capítulo, el **Chief Privacy Officer** (CPO) es el encargado principal de supervisar las prácticas de privacidad y protección de datos dentro de una organización, asegurando el cumplimiento de las obligaciones de privacidad y promoviendo una sólida gobernanza en este ámbito. Su estructura de reporte puede variar, desde el Consejero General hasta departamentos como Legal, Cumplimiento, Seguridad de la Información o Recursos Humanos. Por otro lado, el **Data Protection Officer** (DPO) es un rol independiente designado en cumplimiento con regulaciones específicas, como el RGPD, encargado de supervisar el cumplimiento de normativas de protección de datos, con requisitos particulares de experiencia y ubicación, y su designación puede ser obligatoria en ciertos contextos para garantizar el cumplimiento legal en procesos de tratamiento de datos específicos.

Finalmente, en ocasiones existen otros roles que complementan las funciones del Chief Data Officer (CDO), siendo el **Chief Information Officer** (CIO), **Chief Technology Officer** (CTO) y **Chief Information Security Officer** (CISO).

Al respecto, el CIO se enfoca en la gestión de la información con la tecnología, al armonizar los planes de las entidades con la información, mientas que el CTO se encarga de implementar las tecnologías necesarias para gestionar y analizar estos datos de manera efectiva (OCDE, 2015, 41-44), así como el CISO, al que se le confía la toma de decisiones relacionadas con la seguridad y es responsable de implementar programas de seguridad (Doshi, 2022, 37). Cabe destacar que la Casa Blanca de EE. UU. recientemente ha publicado un memorándum, que establece la hoja de ruta y gestión del uso de IA en la administración pública, obligando la designación de CAIOs (*Chief Artificial Intelligence Officer*) (Executive Office of the President, 2024), dando el surgimiento de un nuevo rol nivel C y que con el tiempo se demostrará su intersección con otros roles.

En conclusión, el papel del *Chief Data Officer* (CDO) ha evolucionado y se ha diversificado para abordar una amplia gama de responsabilidades relacionadas con la gestión y gobernanza de los datos en las organizaciones modernas. Además de los CDO, han surgido roles complementarios como el *Chief Analytics Officer, Chief Data Governance Officer, Chief Data Risk Officer,* y otros más recientes como el *Data Ethics Officer* y el *Data Compliance Officer*, cada uno reflejando la creciente importancia y complejidad del entorno de datos empresariales. La aparición de roles como el *Chief Privacy Officer* y el *Data Protection Officer* resalta la necesidad de asegurar el cumplimiento de regulaciones y normativas de privacidad de datos, mientras que otros roles como el *Chief Information Officer, Chief Technology Officer y Chief Information Security Officer* complementan las funciones del CDO al enfocarse en la gestión tecnológica y la seguridad de la información. En conjunto, estos roles demuestran la creciente centralidad de los datos en las estrategias empresariales y la importancia de una gestión efectiva para aprovechar el valor de los datos mientras se garantiza su integridad, privacidad y seguridad.

Un CDO en una compañía de logística supervisa la implementación de una estrategia de datos para optimizar las operaciones de la cadena de suministro. El CDO lidera la creación de un sistema centralizado para integrar datos provenientes de almacenes, rutas de transporte y clientes. Colabora con el CISO para garantizar que los datos estén protegidos contra brechas de seguridad y con el Chief Analytics Officer CAO para desarrollar modelos predictivos que mejoren la eficiencia de las entregas. Además, establece un comité de ética de datos para revisar el uso de datos anonimizados en análisis avanzados, asegurando el cumplimiento con las regulaciones locales de privacidad.

3.21 Comités, juntas y consejos de datos (Executive Steering Committee, Data Governance Board, Data Stewardship Council)

Finalmente, las estructuras de gobernanza de datos se dividen en tres niveles clave: el **Executive Steering Committee** (Comité Directivo Ejecutivo), el **Data Governance Board** (Junta de Gobernanza de Datos) y el **Data Stewardship Council** (Consejo de Administración de Datos). El Comité Directivo Ejecutivo, compuesto por ejecutivos de alto nivel, impulsa los cambios culturales necesarios y asegura el respaldo y los recursos financieros para la iniciativa; la Junta de Gobernanza de Datos, ubicada en un nivel intermedio, se encarga de establecer prioridades, asignar recursos y garantizar el cumplimiento de políticas y objetivos de gobernanza de datos; y el el Consejo de Administración de Datos, liderado por Stewards de Datos Empresariales, se enfoca en la implementación práctica de la gobernanza de datos, resolviendo problemas cotidianos, estableciendo estándares y procesos, y asegurando la alineación con las necesidades del negocio (Plotkin, 2023, 58-61).

No obstante, algunos autores mencionan que estas estructuras en ocasiones son sinónimos o equivalentes, y que al final son un grupo de partes interesadas representativas que se reúnen regularmente para gobernar y guiar la estrategia del programa de gobierno de datos, con responsabilidades específicas que incluyen revisar y aprobar herramientas de datos, resolver problemas de datos a nivel empresarial, comunicar el valor del programa a la empresa, hacer cumplir varias políticas y aprobar fondos. Además, el CDO puede ser presidente del consejo y se pueden invitar a otras unidades como CIO (Reichental, 2023, 260-261).

En resumen, las estructuras de gobernanza de datos se dividen en tres niveles clave: el Comité Directivo Ejecutivo, la Junta de Gobernanza de Datos y el Consejo de Custodia de Datos. Cada nivel tiene roles específicos, desde liderar el cambio cultural y asegurar recursos financieros hasta establecer prioridades y garantizar la implementación práctica de la gobernanza de datos. Aunque algunas interpretaciones sugieren similitudes entre estas estructuras, cada una tiene responsabilidades únicas dentro del marco de la gobernanza de datos. La participación del Director de Datos y otras unidades, como el Director de Tecnología de la Información, puede mejorar el proceso de toma de decisiones y la

efectividad general de la gobernanza de datos en la empresa. En conjunto, estas estructuras son esenciales para garantizar una gestión efectiva de los datos y su alineación con los objetivos comerciales.

En una empresa de servicios financieros, el Comité Directivo Ejecutivo (*Executive Steering Committee*) aprueba una iniciativa para integrar datos de clientes de distintas regiones, asegurando el presupuesto y alineando el proyecto con los objetivos estratégicos. La Junta de Gobernanza de Datos (*Data Governance Board*) define políticas para estandarizar los datos, establece métricas de calidad y prioriza las áreas críticas de integración. Mientras que el Consejo de Administración de Datos (*Data Stewardship Council*), liderado por *Data Stewards*, implementa las políticas aprobadas resolviendo conflictos de calidad en los datos y definiendo estándares técnicos que garantizan la interoperabilidad entre sistemas.

3.22 Intersección entre los diferentes roles y actores

A continuación, se presenta un ejemplo exhaustivo que considera a todos los actores y roles mencionados en el tratamiento de datos personales dentro de una organización, que en este caso hipotético es un hospital:

Cuadro 4. Ejemplo de gestión de datos en un hospital.

Rol	Descripción	Ejemplo
Interesado (*Data Subject*)	Persona física cuyos datos personales son procesados.	Un paciente, Juan Pérez, acude a un hospital internacional para recibir tratamiento médico. Su información personal, como nombre, edad, historial médico y contacto de emergencia, es recopilada y almacenada en el sistema del hospital.
Responsable del Tratamiento (*Controller*)	Entidad que determina los fines y medios del procesamiento de los datos.	El hospital, como organización, determina los fines y los medios del procesamiento de los datos de Juan Pérez. Su objetivo es usar los datos para brindar atención médica, gestionar citas y cumplir con requisitos legales de salud.

Encargado del Tratamiento (*Processor*)	Entidad que procesa datos en nombre del responsable.	El hospital contrata a una empresa de almacenamiento en la nube para gestionar y almacenar los datos médicos de los pacientes. Esta empresa sigue estrictamente las instrucciones del hospital y no utiliza los datos para otros fines.
Destinatario (*Recipient*)	Entidad que recibe datos personales para un fin específico.	El departamento de oncología del hospital recibe acceso a los datos médicos de Juan Pérez para realizar análisis detallados relacionados con su tratamiento. Además, los resultados del análisis son compartidos con el departamento de facturación para procesar los pagos del seguro médico.
Tercero (*Third Party*)	Entidad que procesa datos bajo acuerdos contractuales sin ser responsable o encargado.	El hospital colabora con una empresa externa para realizar análisis estadísticos de sus datos (anonimizados) con fines de investigación científica. Esta empresa actúa como tercero, manejando los datos bajo estrictos acuerdos contractuales.
Delegado de Protección de Datos (DPO)	Persona responsable de supervisar el cumplimiento normativo en protección de datos.	El DPO del hospital supervisa el cumplimiento normativo del RGPD y otras leyes aplicables. Este delegado capacita al personal sobre prácticas de privacidad, realiza auditorías internas y responde a consultas de los pacientes sobre el uso de sus datos.
Representante (*Representative*)	Entidad designada para interactuar con autoridades regulatorias.	El hospital tiene una sucursal en Europa y designa a un representante local para interactuar con las autoridades regulatorias y los interesados en temas de protección de datos, cumpliendo con el RGPD.
Autoridad de Control (*Supervisory Authority*)	Entidad regulatoria que supervisa el cumplimiento normativo.	La autoridad nacional de protección de datos supervisa las actividades del hospital y asegura que se cumpla con el RGPD. En caso de una violación de datos, el hospital debe notificar a esta autoridad.
Tenedor de Datos (*Data Holder*)	Entidad obligada a gestionar y proteger los datos personales recopilados.	El hospital actúa como tenedor de datos, ya que tiene la obligación de gestionar y poner a disposición los datos médicos de los pacientes de manera segura y conforme a las normativas aplicables.
Usuario de Datos (*Data User*)	Persona que accede y utiliza datos personales de forma legítima.	Un médico especialista del hospital accede legítimamente a los datos de Juan Pérez para planificar su tratamiento. Este usuario debe cumplir con las políticas del hospital y manejar los datos de manera ética.

Productor de Datos (*Data Producer*)	Entidad que genera nueva información o datos derivados.	El historial médico actualizado de Juan Pérez, generado por los dispositivos médicos utilizados durante su tratamiento, convierte al hospital en productor de datos al crear nueva información relevante.
Propietario de los Datos (*Data Owner*)	Persona o rol encargado de establecer políticas de acceso y calidad de los datos.	El director del área de informática del hospital asume el rol de propietario de los datos, siendo responsable de establecer políticas de acceso, supervisar su calidad y garantizar que se utilicen correctamente en el ámbito clínico.
Fideicomisario de Datos (*Data Trustee*)	Entidad que supervisa el acceso a datos en colaboración con otras partes.	El hospital designa un fideicomisario de datos para supervisar el acceso a los datos médicos en colaboración con la empresa de almacenamiento en la nube, asegurando que se cumplan las normativas y se respeten los derechos de los pacientes.
Administrador de Datos (*Data Steward*)	Persona encargada de la calidad, estandarización y resolución de problemas de datos.	El administrador de datos del hospital colabora con los médicos y técnicos para definir estándares de calidad en los datos, resolver problemas de duplicación y garantizar la exactitud de la información.
Custodio de Datos (*Data Custodian*)	Persona o equipo que gestiona la infraestructura técnica para el manejo de datos.	El equipo de TI del hospital actúa como custodio de datos, gestionando las bases de datos, configurando permisos de acceso y estableciendo medidas de seguridad como cifrado y autenticación.
Arquitecto de Datos (*Data Architect*)	Persona que diseña la estructura y flujo de integración de datos en el sistema.	El arquitecto de datos diseña la estructura del sistema que integra datos de varias fuentes (consultas, laboratorios, dispositivos médicos) para facilitar su análisis y uso eficiente en el hospital.
Ingeniero de Datos (*Data Engineer*)	Persona que implementa procesos para recopilar, procesar y almacenar datos.	El ingeniero de datos implementa *pipelines* para recopilar datos de múltiples dispositivos médicos en tiempo real, procesarlos y almacenarlos en un lago de datos centralizado para su análisis posterior.
Científico de Datos (*Data Scientist*)	Persona que analiza datos para desarrollar modelos predictivos o analíticos.	El científico de datos utiliza la información anonimizada para desarrollar modelos predictivos que mejoren la eficiencia de los tratamientos médicos y ayuden a identificar posibles complicaciones futuras.
Analista de Datos (*Data Analyst*)	Persona que genera reportes basados en datos para informar decisiones.	El analista de datos del hospital genera informes sobre tendencias de enfermedades en pacientes de distintas edades y regiones, ayudando a los médicos a tomar decisiones basadas en datos.

Oficiales C-level	Ejecutivos responsables de estrategias y cumplimiento relacionados con los datos.	El Chief Data Officer (CDO) lidera la estrategia de datos del hospital, trabajando con el Chief Information Security Officer (CISO) para garantizar la seguridad de los datos y con el Chief Privacy Officer (CPO) para supervisar el cumplimiento de las normativas.
Comités de Gobernanza de Datos	Grupos encargados de definir políticas, estándares y resolver problemas de datos.	El hospital establece un Comité Directivo Ejecutivo para aprobar políticas de datos, una Junta de Gobernanza de Datos para implementar estándares y un Consejo de Administración de Datos liderado por stewards para resolver problemas operativos.

Cuadro de elaboración propia.

Capítulo 4

ENFOQUES DE MEJORA DE ACCESO, INTERCAMBIO DE DATOS Y SUS TIPOS DE INTERACCIONES.

Capítulo 4. Enfoques de Mejora de Acceso, Intercambio de Datos y sus Tipos de Interacciones

Una vez comprendidos los aspectos legales de los datos, incluyendo sus diversas teorías y clasificaciones, así como la identificación de los actores y sus roles en el ecosistema de datos, el culminar la construcción de un sistema integral basado en datos requiere abordar los enfoques de mejora del acceso, intercambio de datos y las modalidades de interacción entre ellos. Estas etapas finales son fundamentales para garantizar la efectividad, la ética y la maximización del valor de los datos en cualquier contexto, ya sea empresarial, académico o gubernamental. La optimización del acceso y la fluidez en el intercambio de datos, junto con la comprensión de los distintos tipos de interacciones que pueden darse entre los datos y los diversos actores involucrados, son pilares importantes para el desarrollo de un ecosistema de datos robusto y dinámico.

Parar lograr este gran cambio de paradigma sobre los datos, lo primero que hay que equilibrar en el proceso de la legislación sobre los datos radica en su protección y uso, principalmente en cómo regular su recopilación, almacenamiento y utilización, así como los procesos de extracción, análisis y uso, evitando al mismo tiempo la filtración y su abuso de los mismos, y por ello, en este equilibrio, el primer paso es la construcción de un mecanismo de equilibrio dinámico que fomente el uso, pero también su protección eficaz (Yuming, 2021, 1-3).

El sistema de posesión es la base de los derechos reales, mientras que un sistema de intercambio es el núcleo de los derechos de dato (Yuming, 2021, 64). La calidad de los datos determina el valor de los datos, por un lado, se deben recopilar datos de alta calidad; por otro lado, deben estar disponibles datos oportunos, completos y precisos de alta calidad (Yuming, 2021, 227).

Dicho lo anterior, es necesario considerar el desarrollo y uso de datos entre diferentes partes interesadas, como el gobierno, empresas y la sociedad, así como definir claramente las funciones, deberes, actividades y tareas de las diferentes partes, para lograr un equilibrio entre el valor económico y social de los datos gubernamentales y la maximización del valor de los datos personales y

corporativos (Yuming, 2020, 85-86). Los individuos son los productores de datos, el gobierno posee gran cantidad de datos y las empresas son la fuerza principal que promueve el desarrollo y la aplicación de datos. Impulsadas por los intereses de los datos, las empresas invierten grandes cantidades de capital y recursos humanos para desarrollar y mejorar la tecnología de datos; crean nuevos métodos para la producción, recopilación y análisis de datos, buscan mejoras en diversas actividades de datos y enderezan varias relaciones de datos, con el fin de lograr prosperidad, economía y eficiencia en relación con los datos (Yuming, 2020, 88-89).

En concreto, la descentralización de la gestión de derechos de datos, es uno de los avances técnicos más complejos para los derechos sobre los datos (WEF, 2021a, 19-20). Las personas tendrían mayor control sobre sus datos, generando ingresos, mayor certeza y seguridad de que son utilizados de manera adecuada; las empresas podrán adaptar productos existentes, desarrollar nuevos e enriquecer las relaciones con los clientes; los gobiernos podrán crear entornos empresariales positivos de transmisión de datos atractivos para las empresas y que beneficien los derechos de las personas en primer lugar, ayudando a optimizar los servicios prestados y de alta calidad; y para las organizaciones sin fines de lucro, se creará la posibilidad de servir mejor a diversas partes interesadas en función a la información generada a partir de datos del mundo real, pretendiendo construir un mundo en el que los modelos de gobernanza de datos permitan combinar datos de fuentes personales, comerciales y gubernamentales, sin dejar de respetar los derechos (WEF, 2021a, 08-09).

Se necesitan diversos enfoques para mejorar el acceso, intercambio y reutilización de datos a lo largo de todo el ciclo de vida (Organización para la Cooperación y el Desarrollo Económicos (OCDE), 2019a, 24), en su obtención, procesamiento, intercambio, análisis y uso (Verhulst et al., 2015), enfocándose primero, en los diferentes tipos de datos y mecanismos de control de acceso, por ejemplo en diversos datos identificados, seudonimizados, anonimizados, agregados o de dominio personal, privado o público, o control de acceso por las vías de descargas, interfaces de programación de aplicaciones API o cajas de

arena de datos conocidos como *sandboxes*; segundo, en los principales tipos de actores y sus roles, ya sea por medio de usuarios, intermediarios, repositorios, corredores, mercados, sistemas de gestión, almacenes o terceros de confianza de datos; y tercero, en los diferentes enfoques de mejora del acceso y el intercambio de datos y sus tipos relevantes de interacciones, por medio de acuerdos contractuales y mercados de datos (por publicidad, suscripción, tarifas de uso, venta de bienes o servicios, licenciamiento, tarifas de comisión) o por medio de información abierta por parte de los sectores públicos y privados y en el caso de particulares y su ejercicio de datos personales por medio del derecho de autodeterminación informativa y portabilidad de los datos (Organización para la Cooperación y el Desarrollo Económicos (OCDE), 2019a, 25-44).

3.1 Enfoques de Mejora de Acceso

Las fuentes de datos tradicionales pueden ser muy diferentes; desde datos gubernamentales con datos abiertos, datos proporcionados voluntariamente por usuarios, consumidores o empresas, hasta los datos que se generan como *cookies*, son parte fundamental para el acceso a los datos. Sin embargo, estos no son suficientes, ya que el mercado de recopilación, almacenamiento y venta o concesión de licencias de acceso a grandes bases de datos evolucionará y las grandes empresas físicas podrán ingresar a este mercado de competencia con los monstruos corporativos. Además, debe señalarse, que si nadie posee datos, no existe una regla clara sobre si el fabricante del dispositivo, proveedores de red, algoritmos o de servicios en la nube pueden excluirse los datos entre sí, fundamentalmente sin saber quién debería comerciar con quién (Lundqvist, 2018, 192-201).

Por otra parte, Banterle (2018, 439) explica que algunos académicos han propuesto la alternativa de regular el acceso a los datos sin procesar por medio de las leyes en competencia desleal y mediante la introducción en ciertos casos de un sistema de acceso sujeto a tarifa a la información para ser utilizada con fines

comerciales. Bajo una norma de responsabilidad clásica, los propietarios de los datos recibirán una remuneración equitativa y se garantizará el acceso a la información. Esto es fundamental ya que esta propuesta introduce una nueva excepción para la minería de datos, cuyo objeto es equilibrar los derechos de autor con el interés público en acceder a la información.

Existen diversos mecanismos de control de acceso a los datos, conformados por una amplia gama de diferentes mecanismos para acceder a estos. Los más utilizados son acceso a datos mediante i) descargas (*ad-hoc*); ii) interfaces de programación de aplicaciones (API); y iii) las cajas de arena (*sandbox*) de datos, que se reconocen cada vez más como un medio para acceder a datos confidenciales y patentados, paralelamente garantizando los derechos de privacidad y propiedad intelectual de los titulares (OCDE, 2019a, 32).

3.1.1 Descargas

El acceso a datos mediante descargas (*ad-hoc*) implica almacenar los datos, preferiblemente en un formato común y hacerlos disponibles en línea (por ejemplo, a través de un sitio web). Sin embargo, este método plantea varios problemas. La interoperabilidad es un desafío importante para la reutilización de datos entre aplicaciones, incluso cuando se utilizan formatos de máquina legibles comúnmente; estos formatos pueden permitir la portabilidad sintáctica de datos, pero no garantizan la interoperabilidad semántica, que requiere ontologías y metadatos mutuamente entendidos. Además, el acceso a datos mediante descargas puede aumentar los riesgos de seguridad y privacidad digital, ya que una vez descargados, los datos están fuera del control del titular y fuera de su sistema de información, lo que podría resultar en la pérdida de capacidad para hacer cumplir políticas de datos, incluidas aquellas destinadas a proteger la privacidad y los derechos de propiedad intelectual (OCDE, 2019a, 32).

En la práctica, se sigue vendiendo bienes, incluido contenido digital, en el ecosistema de datos, como dispositivos inteligentes con sensores para generar datos y servicios de valor añadido. Esto incluye modelos de ingresos de pago por

descarga, donde los usuarios pagan por cada descarga, ya sea de datos o contenido digital como música y videos (OCDE, 2019a, 40). Sin embargo, si bien hacer que los datos sean legibles y descargables por máquinas es necesario, pero no suficiente, para involucrar a la comunidad de usuarios, incluidos los desarrolladores (OCDE, 2019a, 35).

Indistintamente de lo anterior, las descargas no pueden evitarse y podría ser una buena forma de evitar fallas con sistemas de consultas o de membresía. Un buen proyecto que ha potencializado esta forma de acceso por medio de descargas, es el Portal de Datos Genómicos del Instituto Nacional del Cáncer.

El Portal de Datos Genómicos del Instituto Nacional del Cáncer (NCI GDC) ofrece una forma innovadora de acceder a datos genómicos a través de descargas directas desde el navegador web o utilizando su herramienta de transferencia de datos de alto rendimiento. Esto permite a los investigadores obtener acceso rápido y eficiente a datos sobre el cáncer, incluyendo información clínica, caracterización genómica y análisis de secuencias tumorales de alto nivel, facilitando así el estudio y la comprensión de la biología del cáncer (GDC NCI, s.f.).

La portabilidad de datos y la transferencia de datos personales entre plataformas siguen siendo un tema complejo y desafiante en el ámbito digital. Aunque algunas plataformas como Facebook y Google han facilitado a los usuarios la descarga de sus datos en formatos legibles por máquina, la verdadera interoperabilidad entre estas plataformas aún no se ha logrado completamente. Esta falta de interoperabilidad significa que los usuarios todavía enfrentan obstáculos significativos al intentar transferir sus datos de una plataforma a otra de manera fluida y efectiva. En consecuencia, la descarga de los datos personales de un individuo sigue siendo una opción relevante y necesaria para aquellos que desean ejercer control sobre su información en línea. Sin embargo, es importante que la industria continúe trabajando hacia estándares de interoperabilidad más amplios que permitan a los usuarios transferir sus datos de manera más fácil y segura entre diferentes plataformas, garantizando así una mayor autonomía y control sobre su información personal en el entorno digital.

En resumen, el acceso a datos a través de descargas presenta desafíos significativos en términos de interoperabilidad, seguridad y privacidad. Aunque las descargas son inevitables y a menudo necesarias, especialmente en contextos como el acceso a datos genómicos para la investigación sobre el cáncer, la falta de verdadera interoperabilidad entre plataformas digitales plantea obstáculos para los usuarios que desean controlar y transferir sus datos de manera fluida. Aunque la descarga de datos personales sigue siendo una opción destacable, es fundamental que la industria trabaje hacia estándares más amplios que faciliten la transferencia segura y efectiva de datos entre plataformas, garantizando así una mayor autonomía y control para los usuarios en el entorno digital.

3.1.2 APIs

API (Interfaz de Programación de Aplicaciones) es una interfaz definida en términos de un conjunto de funciones y procedimientos, que permite a un programa acceder a las facilidades dentro de una aplicación, por lo tanto, el uso de tales facilidades permite a los usuarios personalizar la aplicación para sus propios propósitos e integrar la aplicación en un entorno de desarrollo personalizado (Butterfield et al., 2016).

Las API actúan como puertas de entrada esenciales para usuarios comerciales en plataformas, facilitando la interacción entre programas al permitir el acceso y la transferencia de datos de manera estandarizada. Estas API son esenciales para que los desarrolladores y proveedores de servicios externos utilicen las plataformas y sus funcionalidades. Aunque el acceso y la portabilidad de datos se realizan a través de las API, estas están controladas por las plataformas, lo que implica una estandarización que fomenta la interoperabilidad entre plataformas y nubes. Esta estandarización también puede desempeñar un papel importante en el establecimiento de reglas globales para el control del acceso y el uso de datos a nivel mundial (Lundqvist, 2023, 46).

Las API son fundamentales para la interconexión de datos en aplicaciones, permitiendo a los proveedores de servicios compartir recursos digitales a través de Internet. Un ejemplo es el caso de Transport for London (TfL), que utiliza una API unificada para proporcionar datos de transporte tanto en su propio sitio web como a desarrolladores externos. Las APIs facilitan el acceso directo a los datos necesarios para aplicaciones de software, mientras que los titulares de datos pueden controlar su uso mediante restricciones implementadas en las APIs (OCDE, 2019a, 32-33).

Una ventaja clave de una API (en comparación con una descarga de datos ad hoc) es que una API permite que una aplicación de software (o aplicación) utilice directamente los datos que necesita, aunado a que, si los datos están lo suficientemente anonimizados y agregados, esos datos también podrían proporcionarse al público a través de una API que además ayudaría a reducir el nivel de riesgo de reidentificación (OCDE, 2019a, 26-33).

Sin embargo, existen restricciones tecnológicas en el uso de APIs, que son esenciales para los usuarios comerciales de plataformas. Aunque las APIs ofrecen acceso y portabilidad de datos, están controladas por las plataformas, lo que puede resultar en estándares cerrados o abiertos. Estos últimos permiten a terceros construir sistemas interoperables, pero su acceso puede estar restringido por motivos de privacidad, seguridad o técnicos (Lundqvist, 2023, 46-51). Por ejemplo, Twitter utilizó su API para regular el acceso a sus datos, limitando el número de aplicaciones que pueden acceder a ella y restringiendo el desarrollo de aplicaciones que compitan directamente con su plataforma (OCDE, 2019a, 32-33).

En la Unión Europea (UE), las API podrían considerarse protegidas por derechos de autor, mientras que recientemente, en el caso *Oracle vs Google*, la Corte Suprema de los Estados Unidos determinó que las API estaban dentro de la excepción de uso justo; esto podría impulsar llamados para que la UE proteja las API bajo algún tipo de protección de derechos de autor (Lundqvist, 2023, 179).

El modelo de licenciamiento es una estrategia frecuente para generar ingresos a partir de activos intangibles protegidos mediante derechos de propiedad intelectual, como patentes y derechos de autor. Este enfoque se emplea

habitualmente para monetizar datos, software y componentes de software, incluidos algoritmos, bibliotecas y APIs. En este sentido, el licenciamiento de APIs ofrece una forma eficaz de controlar y rentabilizar el acceso a recursos digitales, permitiendo a los titulares establecer condiciones específicas para su uso y obtener ingresos a través de la concesión de licencias de uso.

En conclusión, las APIs son elementos fundamentales en la interconexión y personalización de aplicaciones, permitiendo el acceso y transferencia de datos de manera estandarizada. Si bien ofrecen beneficios significativos en términos de acceso directo a datos y interoperabilidad entre plataformas, también plantean desafíos en cuanto a su control por parte de las plataformas y posibles restricciones tecnológicas. Sin embargo, el debate sobre la protección legal de las APIs, como se observa en el caso Oracle vs Google, sugiere la necesidad de establecer marcos regulatorios claros para su uso y licenciamiento, con el fin de garantizar tanto la innovación como la equidad en su utilización.

3.1.3 Sandbox

Los *sandbox* o cajas de arena de datos son entornos aislados utilizados para acceder y analizar datos, donde los resultados analíticos se exportan solo si no son sensibles. Estos entornos pueden ser creados mediante medios técnicos, como máquinas virtuales aisladas, o mediante presencia física en las instalaciones del poseedor de los datos. Requieren que el código analítico se ejecute en el mismo lugar donde están almacenados los datos. Comparados con otros mecanismos de acceso a datos, los *sandbox* ofrecen un alto nivel de control y son prometedores para proporcionar acceso a datos muy sensibles o propietarios. Ejemplos incluyen el Centro Virtual de Datos de Investigación de Medicare y Medicaid (VRDC) de CMS, que proporciona acceso oportuno a datos de los programas de Medicare y Medicaid, y Flowminder.org, que combina diferentes tipos de datos para apoyar a personas en países de bajos y medianos ingresos, utilizando datos de llamadas móviles de operadores. En ambos casos, se

garantiza la privacidad de los datos y se controla su acceso (OCDE, 2019a, 33-34).

Algunas autoridades como la ICO (s.f.) de UK, han promovido iniciativas como *Regulatory Sandbox*, el cual es un servicio gratuito desarrollado para apoyar a las organizaciones que crean productos y servicios que utilizan datos personales de manera innovadora y segura. Los participantes tienen la oportunidad de interactuar con el equipo del *sandbox* para aprovechar la experiencia más amplia de la ICO y recibir asesoramiento sobre cómo mitigar riesgos y aplicar la protección de datos desde el diseño.

Sin embargo, tanto el EDPB y el EDPS han recomendado aclarar su alcance y objetivos, principalmente con la normativa referente a Inteligencia Artificial de la UE, pues esta debe establecer claramente que la base jurídica de dichos entornos *sandbox* debe cumplir con los requisitos establecidos en el marco de protección de datos existente, expresando preocupaciones sobre cómo se ponderan los intereses de los sujetos de datos y si estos sistemas de IA solo se usarán dentro del *sandbox* (EDPB-EDPS, 2021, 17-18).

Pero al final, los *sandboxes* regulatorios pueden ser útiles para fortalecer la confianza en la economía del intercambio de datos. Estos *sandboxes* permitirían a intermediarios de datos y otros participantes probar nuevos proyectos y tecnologías de intercambio de datos en un entorno seguro y controlado, al mismo tiempo que reciben orientación sobre privacidad (WEF, 2022, 34). E incluso podrían ser considerados una forma de regulación indirecta dinámica, toda vez que ofrecen una vía alternativa para desplegar la IA al permitir la experimentación de nuevas ideas, productos y servicios en entornos seguros sin las restricciones regulatorias habituales, siendo instrumentos valiosos que, junto con otras medidas como certificaciones y evaluaciones de impacto algorítmicas, pueden ayudar a garantizar la seguridad de los usuarios y facilitar la transición hacia sistemas de IA confiables. Al permitir una retroalimentación en tiempo real y una evaluación continua, los *sandboxes* pueden servir como un paso inicial antes de la promulgación de regulaciones definitivas, sin socavar el estado de derecho ni los poderes normativos de los legisladores (Llamas Covarrubias et al., 2022, 47-49).

De hecho, un caso de éxito es el *sandbox* de LunaDNA es un entorno seguro donde los investigadores pueden acceder y analizar datos sin comprometer la privacidad de los miembros. Funciona como un espacio controlado donde los investigadores pueden realizar consultas y análisis sobre los datos almacenados en LunaDNA. Este entorno está diseñado para garantizar que los datos crudos permanezcan dentro de la plataforma y no se compartan externamente. Los investigadores pueden enviar sus consultas o herramientas de análisis al *sandbox*, donde se ejecutan de manera segura. Solo los resultados autorizados se devuelven a los investigadores, como estadísticas, metadatos o modelos de datos, asegurando así la confidencialidad de los datos y la privacidad de los contribuyentes (WEF, 2020a, 40).

En resumen, los *sandbox* de datos representan una innovadora herramienta que proporciona un entorno seguro para acceder y analizar datos sensibles, tanto en el ámbito regulatorio como en el de la investigación. Estos entornos ofrecen un alto nivel de control y privacidad, garantizando que solo se exporten resultados no sensibles. Aunque han surgido preocupaciones sobre su alcance y regulación, se vislumbra su potencial para fortalecer la confianza en el acceso de datos y facilitar el desarrollo de sistemas de inteligencia artificial confiables. Casos exitosos como el *sandbox* de LunaDNA destacan su eficacia en mantener la confidencialidad de los datos mientras se permite la investigación y el análisis. En definitiva, los *sandboxes* emergen como una valiosa herramienta que, junto con otras medidas regulatorias, puede ayudar a promover la seguridad de los usuarios y la transición hacia un entorno de datos más confiable.

3.1.4 Privacy Enhancing Technologies (PETs)

En el pasado, la preocupación por compartir datos estaba marcada por el temor a la exposición y el afán por acumular la mayor cantidad de información posible. Sin embargo, con el tiempo, este paradigma ha evolucionado hacia una economía basada en datos. En este nuevo contexto, las *Privacy Enhancing Technologies* (PETs) desempeñan un papel crucial al permitir desbloquear el potencial del

intercambio de datos. Estas tecnologías no solo facilitan el acceso a la información, sino que también administran la privacidad y seguridad de los datos, lo que resulta fundamental en un entorno donde la confianza y la protección de la información son prioritarias.

Las PETs son soluciones tecnológicas diseñadas para proteger la privacidad en sistemas, aplicaciones o servicios durante su desarrollo. Su implementación sigue estrategias y patrones definidos previamente, si bien su efectividad puede variar con el tiempo debido al rápido cambio tecnológico, estas tecnologías pueden ser herramientas individuales o arquitecturas complejas de sistemas de información (AEPD, 2019, 26).

Y si bien se ha dedicado mucho tiempo a la lucha competitiva para acumular grandes cantidades de datos, se ha prestado menos atención al creciente deseo de las empresas por desbloquear el poder del intercambio de datos. De hecho, se ha avanzado demasiado en términos tecnológicos para poder utilizar datos disponibles, experimentado resultados innovadores y técnicas para administrar la privacidad de los datos, para así permitir a las instituciones desbloquear nuevos valores, los cuales son:

1. Privacidad Diferencial (*Differential Privacy*): Donde se agrega ruido a un conjunto de datos de modo que es imposible aplicar ingeniería inversa a las entradas individuales.
2. Análisis Federado *(Federated Analysis*): Donde las partes comparten los conocimientos del análisis de sus datos sin compartir los datos en sí.
3. Encriptación Homomórfica (*Homomorphic Encryption*): Donde los datos se encriptan antes de compartirlos de manera que se puedan analizar, pero no decodificar en la información original.
4. Pruebas de Conocimiento Cero (*Zero-Knowledge Proofs*): Donde los usuarios pueden demostrar su conocimiento de un valor sin revelar el valor en sí.
5. Computación Segura de Múltiples Partes (*Secure Multiparty Computation*): Donde el análisis de datos se distribuye entre múltiples partes, de modo

que ninguna parte individual pueda ver el conjunto completo de entradas (WEF, 2019d, 4-8).

Respecto al primer punto, el concepto de **privacidad diferencial** busca garantizar que un atacante con conocimiento previo de los datos no pueda aprender nada nuevo sobre un individuo independientemente de su acceso a la base de datos. Aunque estas garantías ideales son imposibles de lograr, se han desarrollado una serie de mecanismos como que añaden ruido a los datos para preservar la privacidad al publicar conjuntos de datos sensibles (Liu & Li, 2021, 283-284). Técnicamente se garantiza un límite probabilístico al aumento de información que puede obtener un atacante al liberar datos; no garantiza una pérdida cero de privacidad ni anonimización total, sino que ofrece una manera de cuantificar y rastrear la pérdida de privacidad de un individuo en un conjunto de datos liberado, es decir, no es una solución mágica aplicable a todos los casos de ciencia de datos ni bloquea toda la información privada de ser aprendida, más bien, es una herramienta que ofrece a los sujetos de datos una negación plausible mientras se realiza el análisis de datos, permitiéndoles decir que pueden ser parte de un grupo con ciertos rasgos, pero que no necesariamente se aplican a ellos específicamente (Jarmul, 2023, 30-38).

Por otra parte, respecto al segundo tipo de PETs, se encuentra el **análisis federado** (también conocido como aprendizaje federado o análisis descentralizado) que permite utilizar múltiples conjuntos de datos en análisis sin que los datos fuente se compartan directamente. En lugar de ello, los poseedores de datos ejecutan cálculos en sus datos locales y solo comparten los resultados, que suelen ser derivados de datos a nivel de grupo o estadísticas resumidas. Esto implica que los datos fuente no se comparten, lo que elimina los desafíos técnicos asociados con la transferencia de conjuntos de datos. Además, proporciona garantías adicionales de privacidad para los poseedores de datos, ya que estos pueden determinar qué cálculos están permitidos en sus datos. Desde una perspectiva técnica, el cifrado de extremo a extremo puede prevenir que terceros adquieran los derivados de los datos. Además, dependiendo del modelo de

confianza, son posibles protecciones adicionales de privacidad, incluidas tecnologías emergentes como la computación segura entre múltiples partes y la privacidad diferencial (Martin et al., 2024, 41).

De acuerdo con el tercer punto de las PETs, el **cifrado homomórfico** permite operaciones en datos cifrados, lo que significa que la información permanece encriptada incluso durante el proceso de cálculo; esto permite que la nube realice operaciones en los datos cifrados sin necesidad de descifrarlos, manteniendo así la privacidad y seguridad de la información. Permite realizar cálculos sobre datos cifrados, lo que facilita la externalización segura de la computación. Además, protege la privacidad de los usuarios al cifrar y externalizar datos en entornos de nube comercial, aplicable a diversos sectores como servicios financieros, comercio minorista, tecnología y salud (Mattsson, 2022, 76-77).

Conforme al quinto punto de las PETs, las **pruebas de conocimiento cero** son una forma de criptografía que permite demostrar la veracidad de una declaración sin revelar ninguna información adicional más allá de la verdad de la declaración misma. Esta técnica mantiene secreta tanto la firma como partes del mensaje firmado, revelando únicamente las partes del mensaje que se han divulgado y que el probador conoce la firma. Para ser considerada una prueba de conocimiento cero, debe cumplir con tres propiedades: completitud, que asegura que si la declaración es verdadera y ambos usuarios siguen las reglas correctamente, el verificador estará convencido sin necesidad de ayuda artificial; solidez, que garantiza que si la declaración es falsa, el verificador no será convencido en ninguna situación; y conocimiento cero, que asegura que el verificador no obtendrá ninguna información adicional en ningún caso (Preukschat & Reed, 2021, 120-121).

Finalmente, respecto al último tipo de PETs abordado en este capítulo, el **cómputo seguro entre partes** (MPC) en criptografía permite a múltiples partes calcular conjuntamente una función de sus entradas secretas sin revelar más que la salida de la función. Aunque no son esquemas de pseudonimización en sí mismos, los protocolos MPC pueden proporcionar medios para implementar pseudonimización sofisticada, como en el caso del protocolo de intersección de

conjuntos privados. Aunque estos protocolos pueden ser más lentos que enfoques más simples, se ha demostrado que son prácticos para grandes conjuntos de datos, permitiendo la computación privada en cuestión de segundos. Recientes avances han mejorado el equilibrio entre los costos de comunicación y computación en estos protocolos (ENISA, 2021, 23-25).

Para ilustrar cómo estas tecnologías pueden mejorar el acceso a los datos y al mismo tiempo proteger la privacidad, es necesario recurrir a ejemplos, que en este caso podría ser un escenario en el ámbito de la salud. Dicho lo anterior, en el ámbito de la investigación médica, es esencial abordar la colaboración entre múltiples hospitales para estudiar enfermedades poco comunes mientras se protegen adecuadamente los datos de los pacientes. En este escenario, se plantea la situación en la que varios hospitales desean unir esfuerzos para comprender mejor una enfermedad poco común. Cada hospital posee datos detallados de sus pacientes, pero compartir esta información directamente plantea desafíos éticos y de privacidad.

En este sentido, una estrategia para enfrentar esta preocupación es la aplicación de PETs. Estas técnicas agregan ruido a los conjuntos de datos, garantizando que, al compartir estadísticas o resultados de análisis, la identidad de los pacientes permanezca protegida. Por ejemplo, al examinar la prevalencia de la enfermedad o la efectividad de ciertos tratamientos, se evita que los individuos sean identificados.

Otra alternativa es el análisis federado, donde cada hospital realiza análisis en sus propios datos y comparte solo los resultados agregados. Esto permite colaborar en la identificación de patrones de enfermedades o en la evaluación de tratamientos sin compartir datos personales.

Previo al intercambio de datos, se puede recurrir a la encriptación homomórfica. Esta técnica cifra los datos, permitiendo que los investigadores realicen cálculos en ellos sin descifrar la información subyacente, lo que garantiza la confidencialidad.

Las pruebas de conocimiento cero resultan útiles para verificar afirmaciones sobre datos sin revelar detalles sensibles. Por ejemplo, se puede demostrar la eficacia de ciertos tratamientos sin revelar la información médica individual.

Por último, el uso de protocolos de computación segura entre múltiples partes permite realizar análisis complejos que requieren datos de múltiples hospitales. Esto permite calcular funciones de los conjuntos de datos sin revelar información más allá de los resultados, lo que resulta crucial para colaboraciones a gran escala. En resumen, estas tecnologías no solo facilitan la colaboración entre hospitales para investigar enfermedades, sino que también garantizan la confidencialidad y seguridad de los datos de los pacientes en todo momento, promoviendo así avances significativos en el campo de la medicina.

No obstante, sin menoscabo del ejemplo anterior, se debe hacer hincapié, en que lo que es seguro actualmente, es probable que en el futuro no lo sea, y en este sentido las PETs son una gran herramienta de seguridad y privacidad, pero esto no significa dejar de cumplir con regulaciones o cobijarse sin rumbo en la tecnología sin un contexto visionario. Dicho lo anterior, las PETs debe ajustarse a cada entidad y proceso en concreto.

Al respecto, algunos consejos son que, si bien los criterios de evaluación de PETs no tienen un método único, hay puntos de partida claros. Primero, se deben considerar los criterios del caso de uso, que incluyen requisitos específicos y preferencias del usuario. Segundo, se debe recopilar información de equipos externos, como el equipo de privacidad y de seguridad de la información. Tercero, se realiza modelado de amenazas y riesgos de privacidad para identificar y priorizar los riesgos que deben abordarse. Cuarto, se identifican los requisitos técnicos de privacidad y se incorporan aprendizajes previos. Finalmente, se determinan las tecnologías factibles, se experimenta y se implementa, adaptándose a los requisitos del producto y del caso de uso, y manteniendo una comunicación constante con todas las partes interesadas (Jarmul, 2023, 251-252).

En resumen, el paradigma del intercambio de datos ha evolucionado de la acumulación masiva a una economía basada en datos, donde las PETs desempeñan un papel crucial al gestionar la privacidad y seguridad de la

información. Estas tecnologías, como la Privacidad Diferencial, el Análisis Federado, la Encriptación Homomórfica, las Pruebas de Conocimiento Cero y la Computación Segura entre Partes, ofrecen soluciones tecnológicas para preservar la privacidad mientras se permite el intercambio y análisis de datos. Si bien no son soluciones mágicas y su efectividad puede variar, representan herramientas fundamentales en un entorno donde la confianza y la protección de la información son primordiales, abriendo nuevas oportunidades para la innovación y el desarrollo tecnológico.

3.2 Intercambios de datos

Una vez que se han abordado con los enfoques de mejora de acceso a los datos mediante descargas, APIs, *sandbox* y *Privacy Enhancing* Technologies (PETs), es necesario continuar la investigación y ahora abordar el intercambio de datos. Si bien estos enfoques han sentado las bases para la seguridad, privacidad y accesibilidad de los datos, el intercambio efectivo de información entre diferentes entidades y sistemas es esencial para impulsar la colaboración, la innovación y el progreso en diversos campos. Este paso hacia el intercambio de datos permite aprovechar plenamente el potencial de la información disponible, fomentando así el avance en la investigación, el desarrollo tecnológico y la toma de decisiones informadas.

El primer punto medular para desarrollar una comprensión real sobre el intercambio e interacciones con los datos radica en tener en cuenta que los datos no son rivales ni fungibles. Es decir, la rivalidad significa que si bien el capital físico no puede ser utilizado por varias personal al mismo tiempo, el capital de datos puede ser utilizado por varias personas al mismo tiempo y por un número ilimitado de partes debido a su reproducibilidad, y en este sentido, ser no fungible significa que el capital físico se puede reemplazar, pero los datos no, y en un ejemplo, se puede sustituir un barril por otro, pero un dato no puede sustituir a otro, porque cada dato tiene información y valores diferentes (Yuming, 2021, 15-17).

En el intercambio de datos todas las partes ganan, pues, aunque la razón sea intercambiar datos por datos o datos por dinero, al final se desbloquea el valor de los mismos (Treder, 2019, 30). Empero, desde una perspectiva económica, el acceso a los datos presenta desafíos y oportunidades, pues si bien la definición clara de los derechos de propiedad y el fomento del comercio de datos podrían resolver el problema, enfrentamos obstáculos legales y económicos significativos. Marcos legales actuales, como el RGPD, otorgan ciertos derechos a los sujetos de los datos, pero persisten ambigüedades en la asignación de derechos entre sujetos y recolectores. La falta de mercados de datos multilaterales y la dificultad para establecer la calidad de los datos antes de su acceso son barreras adicionales (World Bank Group, 2021, 32).

Pero ¿Quiénes son los intermediarios de los datos? Una respuesta breve sería que son servicios de intermediación que incluyen conjuntos de datos establecidos conjuntamente por varias personas físicas o jurídicas con intención de otorgar una licencia para el uso de dichos datos a las partes interesadas, de forma que, todos los participantes que contribuyan al conjunto de datos reciban una recompensa por su contribución al conjunto; es decir, son organizaciones que permiten el acceso a datos, su control y la capacidad de dirigir sus datos en apoyo a organizaciones, empresas o grupos con los que estén alineados (Thompson, 2023, 115-116).

Al respecto, algunos intermediarios en el intercambio de datos son los corredores de datos (*data brokers*), mercados de datos (*data markets*), grupos de datos (*data pools*), datos comunes y clubes de datos (*data commons and data clubes*), cooperativas de datos (*data cooperatives*) y fideicomisos de datos (*data trust*) (Banco Mundial, 2021, 277-278). Y en el mismo sentido, algunos participantes son las entidades gubernamentales, sociedad civil e individuos, academia, sector privado y las organizaciones internacionales y regionales (Banco Mundial, 2021, 306-307).

En el presente capítulo, se explorarán una variedad de conceptos relacionados con los intercambios de datos en el ámbito tecnológico y regulatorio. Desde las transferencias de datos hasta los mercados de datos, pasando por los

corredores de datos, los intercambios de datos, los espacios de datos y los ecosistemas de datos, se examinará cómo se estructuran y operan estos diferentes modelos. También se discutirán conceptos como los cultivos de datos, las refinerías de datos, los grupos de datos, los datos comunes y los clubes de datos, así como las cooperativas de datos, las colaboraciones de datos y los fideicomisos de datos. Este análisis nos permitirá comprender mejor las dinámicas, los desafíos y las oportunidades que rodean al intercambio de datos en la era digital.

Previo a comenzar el análisis, a continuación, se comparte una tabla comparativa de las diferentes formas de intercambio de datos.

Cuadro 5. Aspecto, definición y propósito principal de intercambios de datos.

Aspecto	Definición	Propósito Principal
Transferencias de Datos	Movimiento y almacenamiento de datos entre fronteras	Facilitar el flujo de datos entre fronteras
Mercados de Datos	Espacios donde se valora y negocia la información	Valorar y comercializar datos
Corredores de Datos	Intermediarios que recopilan y venden datos	Recopilar y vender datos para uso comercial
Intercambios de Datos	Plataformas que facilitan el intercambio y monetización de datos	Facilitar el acceso y la ejecución de algoritmos sobre datos
Espacios de Datos	Permiten gestionar todas las fuentes de datos de una organización	Modelar las relaciones entre las distintas fuentes de datos y gestionar su acceso.
Ecosistemas de Datos	Permiten extraer valor de las cadenas de valor de los datos, facilitando la colaboración entre diversas organizaciones e individuos	Permitir extraer valor de las cadenas de valor de los datos, facilitando la colaboración entre diversas organizaciones e individuos
Cultivos de Datos	Datos seleccionados y desarrollados activamente para su valor comercial	Seleccionar y desarrollar activamente datos para su valor comercial
Refinerías de Datos	Procesan datos poliestructurados dispares en formatos utilizables para análisis	Procesar y transformar datos poliestructurados en formatos utilizables para análisis

Grupos de Datos (Data Pools)	Herramientas que permiten la agregación y compartición de datos entre empresas para obtener perspectivas e información mutuas	Combina datos en un único y amplio conjunto con el objetivo de obtener perspectivas e información que de otra manera no podrían obtener de manera independiente.
Uniones de Datos (Data Unions)	Estructuras organizativas que permiten a los individuos agrupar y compartir sus datos para su monetización de manera colectiva y justa	Permitir a los individuos agrupar y compartir sus datos para su monetización de manera colectiva y justa.
Datos Comunes y Clubes de Datos (Data Commons and Data Clubs)	Redes de relaciones entre titulares de derechos de datos que tienen igual derecho a un recurso de datos común e indivisible.	Red de relaciones entre titulares de derechos de datos que tienen igual derecho a un recurso de datos común e indivisible.
Cooperativas de Datos (Data Cooperatives)	Red de acuerdos entre pares con intereses mutuos que permite la agrupación de recursos de datos.	Asociación público-privada que permite la colaboración y el intercambio de información entre sectores y actores diversos.
Colaboraciones de Datos (Data Collaboratives)	Relación de intercambio de datos que restringe el acceso y uso a los miembros de la colaboración, utilizado en proyectos que benefician el desarrollo de políticas públicas y empoderan a los participantes.	Modelo alternativo frente al enfoque predominante basado en el mercado para la gobernanza de datos, promoviendo la colaboración entre entidades públicas y privadas para el beneficio común.
Fideicomisos de Datos (Data Trust)	Solución legal y no tecnológica que establece un marco jurídico para el intercambio de datos, conceptualizando los datos como propiedad y delegando su control a un fiduciario.	Promover la confianza entre partes, asegurar el cumplimiento de objetivos comunes y garantizar el uso ético de los datos.

Tabla de elaboración propia.

3.2.1 Transferencias de Datos (Data Transfers)

La capacidad de trasladar, almacenar y procesar datos fuera de las fronteras nacionales es fundamental para la economía internacional de datos, pues la política de flujos de datos transfronterizos ha cobrado una gran importancia. Pero

el aumento de leyes y políticas contemporáneas actúan como barreras para este tipo de intercambio internacional de datos (WEF, 2020b, 05-15). Ante esto, se pueden identificar diversas taxonomías que clasifican la regulación de transferencia de datos transfronterizas en todo el mundo en 5 arquetipos base:

1. Régimen de flujo no incondicional**:** los datos pueden fluir libremente a través de las fronteras sin requisitos específicos, por ejemplo, el Reglamento de datos no personales de la UE.
2. Régimen de flujo condicional**:** Los datos no se pueden transferir al extranjero a menos que el país receptor, el controlador de datos y/o el procesador de datos cumplan ciertas condiciones. Hasta la fecha, este régimen se ha aplicado normalmente a los datos personales, por ejemplo, el RGPD.
3. Requisito de almacenamiento local**:** Ciertos datos no se pueden transferir a través de fronteras a menos que se almacene una copia dentro de los límites de la jurisdicción. Este tipo generalmente se aplica a ciertos registros fiscales y contables, documentos corporativos, archivos públicos y datos de usuarios en poder de empresas de telecomunicaciones u otros intermediarios de Internet. Por ejemplo, Vietnam en su Ley de Ciberseguridad.
4. Requisito de procesamiento local**:** El procesamiento principal de los datos debe realizarse en centros de datos ubicados en el país implementador. La empresa debe construir o arrendar un centro de datos en el país, o cambiar a proveedores locales de soluciones de procesamiento de datos. En ocasiones, este enfoque se adopta con respecto a los datos en poder de las autoridades públicas o en sectores sensibles como las finanzas, la salud y las telecomunicaciones. Por ejemplo, Rusia en su Ley de Protección de Datos.
5. Prohibición de la transferencia de datos**:** Los datos se deben almacenar, procesar y acceder a estos dentro del territorio del país implementador. Este tipo se diferencia del procesamiento local en que a la empresa no se le permite ni siquiera enviar una copia de sus datos al exterior. Suele aplicarse

a datos clasificados como especialmente sensibles, como datos sanitarios o datos financieros. Por ejemplo, Australia en su Ley de Registros Médicos (WEF, 2019c, 8-12).

Otro punto es, que desde una visión universal, los flujos de datos transfronterizos son esenciales para la economía digital, ya que impulsan industrias emergentes, sin embargo, los gobiernos adoptan cada vez más políticas que restringen o prohíben los flujos a sus limitaciones territoriales, indicando la falta de confianza entre gobiernos y la creencia de que los datos fuera de las fronteras locales no pueden cumplir con los objetivos de las políticas nacionales, privacidad, seguridad y acceso a los datos. Ante esto, el Foro Económico Mundial hizo un llamado a construir un orden internacional para los flujos de datos que involucren a los principales expertos, empresas y partes interesadas para convertir un concepto histórico en una arquitectura de gobernanza. Se tomó como referencia la visión del ex primer ministro japonés Shinzo Abe con su idea de *Data Free Flow with Trust (DFFT)*, donde la apertura de los flujos coexiste con la confianza de que los datos pueden estar seguros a través de las fronteras (WEF, 2021d, 20).

Cabe destacar que la visión de *Data Free Flow with Trust (DFFT)* fue establecida en la Cumbre del Grupo de los 20 (G20) celebrada en Osaka Japón en junio 2019, donde líderes de todo el mundo discutieron cómo lidiar con la economía digital en rápida expansión. Dicho marco llamado *Osaka Track* o declaración de *Osaka* pretendió crear reglas internacionales sobre el uso de los datos y estipulaba la idea de *DFFT* para ayudar la economía digital, como un marco para la distribución de datos respaldados por la confianza (WEF, 2019a, 07). Esta declaración contemplaba una sección titulada Innovación: digitalización, libre flujo de datos con confianza (artículos 10 a 12), refiriéndose a los desafíos de la privacidad de los datos en el contexto de la protección de datos, propiedad intelectual y ciberseguridad para facilitar un libre flujo en fortalecimiento al consumidor y confianza empresarial, necesitando marcos legales e internacionales (G20, 2019).

Pese al esfuerzo y suscripción por 24 países incluidos EE. UU., China, Rusia, la UE, países de América Latina y del Este de Asia, cuatros países no participaron, siendo India, Egipto, Indonesia y Sudáfrica, con el argumento de que los países en desarrollo necesitan tiempo y políticas para construir una comprensión más profunda del tema y formular marcos regulatorios locales antes de participar de manera significativa en negociaciones de comercio electrónico (Greenleaf, 2019, 2). De hecho, India ha abogado por un enfoque proteccionista con el objetivo de salvaguardar a su país del "colonialismo de datos", es decir, evitar que los países ricos obtengan beneficios de los flujos de datos transfronterizos a costa de dañar los intereses nacionales (UNCTAD, 2021, 110-111).

La circulación de datos es esencialmente un permiso para usar los datos, pudiendo ser, permiso uno a uno, permiso uno a muchos y permisos mutuos; la primera característica básica del intercambio de datos es que los sujetos se limitan a un ámbito específico y debe haber al menos dos sujetos; el segundo es que los sujetos específicos tienen permiso para usar los datos que son propiedad o están controlados por todos los sujetos dentro de un cierto alcance basado en un mecanismo de permiso mutuo de uso (Yuming, 2021, 118-119).

Empero, las transacciones de datos son objeto de una relación jurídica con la finalidad de ser comercializados, sin embargo, los datos que involucren (1) Datos que involucran seguridad nacional, seguridad pública y privacidad personal; (2) Datos que involucren secretos comerciales sin la autorización y consentimiento del titular legal del derecho; (3) Datos que involucren información personal sin el consentimiento explícito del sujeto de la información personal; datos que involucren la información personal del menor mayor de 14 años sin el consentimiento expreso del menor o de su tutor; datos que involucren la información personal del menor de 14 años sin el consentimiento explícito de su tutor; (4) Datos obtenidos mediante fraude, engaño, declaraciones engañosas o de canales ilegales; (5) Datos expresamente prohibidos por otras leyes, reglamentos o acuerdos legales, no pueden ser comercializados (Yuming, 2021, 123).

Por ejemplo, un experto en ciberseguridad que trabaja para una empresa internacional que desarrolla *software* y necesita transferir datos personales de clientes desde la UE a EE. UU., debe garantizar el cumplimiento del Régimen de flujo condicional estipulado por el RGPD. Para ello, verifica si EE. UU. cuenta con un nivel adecuado de protección de datos o implementa cláusulas contractuales estándar (SCCs) con las partes receptoras, adopta medidas complementarias como cifrado o anonimización para mitigar riesgos y documenta todo el proceso, asegurando así que las transferencias internacionales sean legales, seguras y conformes con las regulaciones aplicables.

3.2.2 Mercados de Datos (Data Markets)

Valorar los datos sigue siendo un desafío, y la cadena de valor de los datos es clave para su estimación. El valor surge del proceso de transformación de datos crudos en inteligencia digital, que puede ser monetizada o utilizada para fines sociales. Aunque los datos individuales carecen de valor, su agregación y procesamiento son fundamentales. No hay mercados desarrollados de datos crudos; la mayoría de las veces se habla de mercados de inteligencia digital. Las estimaciones del valor de los datos se centran en estos mercados, pero ofrecen poca información sobre el valor de los datos crudos. Los mercados de datos han crecido significativamente en los últimos años, destacando la posición dominante de Estados Unidos en este ámbito (UNCTAD, 2021, 17-18).

Hoy en día, diversas organizaciones e individuos adquieren datos personales sensibles, impulsando el mercado de datos. Actores clave incluyen agencias de publicidad, medios, empresas de análisis de datos, software y corredores de datos. Es importante entender cómo se utilizan estos datos y por qué son valiosos para evaluar el riesgo al almacenar, procesar o transferir conjuntos de datos (Davidof, 2020, 34).

La competencia no regulada en mercados basados en datos tiende a fallar bajo monopolios fuertes. Esto no estimula la creatividad ni beneficia la innovación. Además, la falta de mercados de datos impide la transferencia de conocimientos

hacia empresas que podrían utilizarlos para investigación y desarrollo. Las plataformas con grandes cantidades de datos relevantes compiten con empresas físicas, desincentivando así la inversión en I+D por parte de estas últimas, que a menudo se convierten en subcontratistas de las plataformas (Lundqvist, 2023, 64).

Comprender los mercados de datos implica ir más allá de la definición económica tradicional de mercado y considerar cómo las plataformas digitales y las refinerías de datos están transformando el intercambio de información. Estos mercados involucran la valoración y el intercambio de tramos de datos derivados de perfiles de comportamiento, con una estructura que requiere objetos valorables, agencias de cálculo distribuidas y una estructura institucional comúnmente entendida. Es esencial reconocer que los bienes en estos mercados no son los datos en sí mismos, sino el acceso a grupos específicos de datos, y que las plataformas digitales juegan un papel central en la organización y funcionamiento de estos intercambios (Cohen, 2019, 68-70).

Los mercados de datos, tanto nacionales como internacionales, mayormente consideran los datos como activos privados gobernados por derechos individuales. Esto se basa en la premisa de que los individuos son la unidad relevante para los derechos de datos, lo que los convierte en capital y mercancía comerciable. Economistas los ven como recursos no rivales y no excluibles, fomentando un mercado libre donde los datos son extraídos y comercializados, principalmente beneficiando a quienes pueden recopilarlos y procesarlos (Solano et al., 2022, 21-22).

En síntesis, los datos son un recurso valioso que se transforma en inteligencia digital, lo que puede generar beneficios económicos o sociales. Aunque los datos crudos por sí solos carecen de valor, su agregación y procesamiento son fundamentales. Los mercados de datos han crecido considerablemente, especialmente en Estados Unidos, donde diversos actores como agencias de publicidad, medios y empresas de análisis de datos adquieren datos personales sensibles. Sin embargo, la falta de regulación puede conducir a monopolios que obstaculizan la innovación y la transferencia de conocimientos. Estos mercados no se centran en los datos en sí, sino en el acceso a ellos, con

plataformas digitales que juegan un papel crucial en su organización y funcionamiento.

Al respecto, para luchar en contra de estos monopolios, la UE ha emitido la Ley de Mercados Digitales y la Ley de Servicios Digitales. La Regulación de Mercados Digitales (Digital Markets Act) busca asegurar un sector digital justo y competitivo, promoviendo la innovación y protegiendo a los usuarios en línea; establece reglas para su conducta, incluyendo la promoción de la interoperabilidad y el acceso equitativo a los datos. Mientras que, la Regulación de Servicios Digitales (Digital Services Act) busca crear un entorno en línea más seguro en la Unión Europea, protegiendo los derechos de los consumidores, definiendo responsabilidades para las plataformas en línea y abordando el contenido ilegal. Introduce normas para proveedores de servicios intermediarios y plataformas muy grandes, promueve la transparencia y empodera a los usuarios.

Por otra parte, el Reglamento de Gobernanza de Datos (2022), busca regular los *data markets*, plataformas donde las empresas pueden compartir datos entre sí. Se excluyen servicios como almacenamiento en la nube o análisis de datos, siempre y cuando solo proporcionen herramientas técnicas para compartir datos sin establecer relaciones comerciales. Ejemplos de servicios cubiertos incluyen mercados de datos y sistemas de intercambio de datos abiertos. Y de manera complementaria el Reglamento de Datos (2023), menciona que la alfabetización de datos es crucial para que usuarios y empresas comprendan el valor de los datos y participen en un mercado justo e inclusivo. Se necesita promover la educación en este ámbito para mejorar las condiciones laborales y favorecer la innovación en el mercado de datos, asegurando que los usuarios comprendan sus derechos y obligaciones en este entorno.

Por ejemplo, un experto en mercados de datos podría trabajar con una empresa tecnológica que desea monetizar datos generados por dispositivos IoT en hogares inteligentes. Este experto primero identifica datos como patrones de consumo energético, y los agrega para crear inteligencia digital útil. Luego, negocia con plataformas de mercados de datos para ofrecer estos conjuntos de datos procesados a empresas de energía interesadas en desarrollar modelos

predictivos o personalizar ofertas para clientes. Durante este proceso, el experto asegura el cumplimiento normativo, como las disposiciones del Reglamento de Gobernanza de Datos de la UE, y promueve transparencia y protección de los derechos de los usuarios involucrados.

3.2.3 Corredores de Datos (Data brokers)

Los intermediarios de datos han desempeñado un papel esencial en muchos mercados, actuando como nodos que conectan a compradores minoristas con vendedores en diversos entornos, desde la Bolsa de Nueva York hasta los mercados de pescado. Su función como intermediarios también podría aplicarse a los mercados de datos. Los intermediarios de datos podrían analizar las necesidades potenciales de los consumidores de datos, proponer los mejores conjuntos de datos y recopilar información. Podrían evaluar y asumir riesgos en nombre de los consumidores de datos. Se espera que los proveedores de servicios de gestión de datos coordinen estrechamente con los intermediarios para desarrollar mercados de datos y supervisarlos para garantizar que sean confiables para todas las partes (WEF, 2021e).

Es aquí cuando surgen los *Data Brokers*, los cuales constituyen un sector privado con motivaciones muy diferentes a las cuestiones académicas en cuanto a la producción y reutilización de datos. Mientras que existen infraestructuras de datos de investigación que buscan cumplir un bien público creando un "común de datos" o *data commons* que facilite y promueva el intercambio y reutilización de datos, los data brokers (a veces llamados agregadores, consolidadores o revendedores de datos) capturan, reúnen y empaquetan datos en infraestructuras de datos privadas para su alquiler (para uso único o bajo condiciones de licencia) o reventa con fines de lucro (Kitchin, 2014, 44).

Los productos incluyen listas de clientes potenciales que cumplen ciertos criterios (incluidos detalles como nombres, direcciones, números de teléfono, direcciones de correo electrónico, así como información como género, edad, etnia, presencia de niños, ingresos, valor del hogar, propiedad de tarjetas de crédito,

estado crediticio, patrones de compra y pasatiempos), verificaciones de antecedentes, una serie de productos de datos derivados, donde los *brokers* agregan valor a través de la integración, y productos de análisis de datos que, en su mayoría, se utilizan para microsegmentar campañas publicitarias y de marketing, evaluar la solvencia crediticia y clasificar socialmente a individuos, proporcionar servicios de rastreo y suministrar análisis detallados de negocios (Kitchin, 2014, 45-47).

Por encima de abordar el debate de la calidad de los datos, puesto que estos datos son secundarios, cuestionando su responsabilidad respecto de la calidad de la información (Fleckenstein & Fellows, 2018, 8). Existe un problema que va más allá de la calidad, siendo problemas de transparencia, lealtad, información y auditabilidad.

Como indica Pasquale (2015), sin transparencia en el origen y destino de los datos, es imposible medir el alcance del uso indebido de la información (29), y en consecuencia, los *data brokers* pueden esconder prácticas cuestionables, ya que podrían vender información personal y a menudo, los consumidores desconocen cómo sus datos son utilizados y cómo esto puede afectar sus vidas (19-21), incluyendo la creación de perfiles o *profiling* (32). Además, estos *data brokers* pueden obligar contractualmente a los clientes a no revelarlos como fuentes (26).

En este sentido, los *data brokers* deberían ser transparentes respecto a qué datos acumulan, comercian y venden, y por otra parte, los usuarios deberían tener el derecho a inspeccionar, corregir y disputar datos inexactos, así como saber la fuente de los datos, salvo que existan excepciones legítimas para mantener su anonimato; irónicamente algunos intermediarios podrían negarse a revelar sus fuentes por acuerdos de confidencialidad con sus clientes. Además, los reguladores y autoridades de privacidad deberían exigir a los auditores que comprendan a fondo las prácticas de los intermediarios de datos para detectar y prevenir rápidamente incumplimientos en estándares de recopilación, etiquetado y filtrado de datos (Pasquale, 2015, 145-151).

La evolución de la industria de los intermediarios de datos ha pasado por varias generaciones. La primera generación incluyó empresas como InfoUSA, centradas principalmente en la industria del correo directo; la segunda generación incluyó empresas como Experian, Acxiom y Epsilon, destacadas en el ámbito de las puntuaciones de crédito y la determinación de la solvencia crediticia; hoy en día, tenemos una tercera generación de intermediarios de datos centrados en recopilar información de actividades en línea, como datos de ubicación, comportamiento de navegación y compras (Thompson, 2023, 7).

Finalmente, cabe hacer la aclaración, que los intermediarios de datos (***Data brokers***) también se conocen como proveedores de datos (***data providers / suppliers***), o intermediarios de información (Hollifield & Cofey, 2023, 154).

En conclusión, el papel de los *data brokers* en los mercados contemporáneos es significativo y multifacético. Si bien pueden facilitar eficazmente la conexión entre compradores y vendedores de datos, especialmente en un mundo cada vez más digitalizado, también plantean desafíos éticos y de transparencia. A medida que la industria evoluciona a través de diferentes generaciones de intermediarios, desde aquellos centrados en datos offline hasta los más modernos que se enfocan en la recopilación de datos en línea, se hace evidente la necesidad de abordar cuestiones relacionadas con la privacidad, la calidad de los datos y la responsabilidad. Es esencial establecer regulaciones claras y mecanismos de supervisión para garantizar que estos intermediarios operen de manera transparente y ética, protegiendo los derechos y la privacidad de los individuos mientras fomentan la innovación y el intercambio de datos de manera responsable.

Un profesional en corredores de datos puede colaborar con empresas de marketing digital para adquirir datos específicos que optimicen campañas publicitarias. Este experto evalúa la calidad, transparencia y origen de los datos, asegurando el cumplimiento normativo, como el RGPD, y negociando contratos que incluyan cláusulas de auditoría y corrección, minimizando riesgos legales y éticos mientras maximiza la utilidad de la información.

3.2.4 Intercambios de Datos (Data Exchanges)

Los intercambios de datos han emergido como un elemento clave para la economía de los datos, creando espacios donde facilitan que cualquier empresa monetice sus activos de datos y cree nuevas fuentes de ingresos. Existen principalmente 3 tipos de intercambios de datos: primero, Intercambio de datos entre pares, donde los datos pueden ser compartidos directamente, por ejemplo, entre dos compañías; segundo, Intercambio de datos privados, como un consorcio industrial que comparte datos; y tercero, mercado de datos, siendo un intermediario de acceso público para cualquier compañía que requiera datos sobre consumidores (Thompson, 2023, 113-115).

Las plataformas de intercambio de datos son plataformas que recopilan datos de diversas fuentes y permiten a terceros ejecutar algoritmos sobre estos datos. Como resultado, estos terceros pueden generar ideas (conocimiento) con nuevas fuentes de datos. Por lo tanto, los intercambios de datos dan lugar al concepto de datos compartidos, que es el siguiente paso natural del big data. Estos intercambios de datos van a transformar profundamente la forma en que se genera el conocimiento, abriendo el horizonte para el próximo nivel de generación de valor basado en datos (Parra-Moyano et al., 2021, 35-36).

Cabe hacer la aclaración que los *data exchanges* y los *data markets* son dos conceptos clave en el ámbito de la gestión de datos. Mientras que los data exchanges se enfocan en facilitar el intercambio de datos entre diferentes partes, actuando como intermediarios para estas transacciones, los data markets se centran en la compra y venta de datos como productos o servicios. En un *data exchange*, los participantes pueden intercambiar datos entre sí según acuerdos establecidos, mientras que en un *data market*, los datos se ofrecen para su compra a través de transacciones monetarias. Aunque ambos conceptos tienen el objetivo de facilitar el acceso a los datos, difieren en la naturaleza de las transacciones que facilitan, con los *data exchanges* priorizando el intercambio y los *data markets* la compra y venta de datos estructurados para su fácil consumo por parte de los compradores (WEF, 2021e, 03).

De manera formal, los considerandos 3, 32 y 33 del Reglamento de Gobernanza de Datos (2022), establecen que, para incrementar la confianza en los servicios de intercambio de datos, es esencial establecer un marco regulatorio a nivel de la Unión Europea que garantice requisitos armonizados para una prestación confiable de dichos servicios. Esto permitirá que los titulares de datos y los usuarios tengan un mayor control sobre el acceso y uso de la información, conforme a la legislación vigente. Además, se debe promover la neutralidad de los intermediarios de datos para asegurar un entorno competitivo y evitar conflictos de interés, así como facilitar la interoperabilidad entre diferentes plataformas.

Y por su parte, el artículo 22 el Reglamento de Datos (2023) destaca la cooperación y asistencia mutua entre organismos públicos y entidades de la Unión Europea para implementar este capítulo de manera consistente. Se establece que los intercambios de datos en este contexto deben cumplir con el propósito para el cual fueron solicitados y no pueden ser utilizados de manera incompatible. Antes de solicitar datos a otro Estado miembro, se debe notificar a la autoridad competente, y esta debe examinar la solicitud en función de los requisitos establecidos. Los organismos solicitantes deben considerar el consejo de la autoridad competente antes de tomar cualquier otra acción, como volver a enviar la solicitud.

En resumen, los intercambios de datos, que incluyen los *data exchanges* representan una evolución crucial en la economía de los datos. Facilitan la monetización de los activos de datos y la creación de nuevas fuentes de ingresos para las empresas. Estos intercambios, divididos en tres tipos principales, están transformando la forma en que se genera conocimiento, al permitir el acceso a datos compartidos y la ejecución de algoritmos sobre ellos. Mientras que los *data exchanges* se centran en facilitar el intercambio de datos entre partes, actuando como intermediarios, los *data markets* se enfocan en la compra y venta de datos como productos o servicios. Ambos conceptos tienen como objetivo primordial ampliar el acceso a los datos, aunque difieren en la naturaleza de las transacciones que facilitan, con los *data exchanges* priorizando el intercambio directo y los *data markets* enfocándose en transacciones monetarias para datos

estructurados. Esta evolución marca el próximo nivel en la generación de valor basado en datos, promoviendo una mayor colaboración y aprovechamiento de recursos digitales.

Por ejemplo, un profesional en intercambios de datos colabora con un consorcio industrial de energías renovables para crear una plataforma privada y segura que comparte datos anonimizados, como consumo energético y rendimiento de equipos. Esto garantiza confidencialidad y cumplimiento normativo, permitiendo a las empresas generar modelos predictivos, optimizar recursos y desarrollar soluciones innovadoras, promoviendo colaboración y valor compartido.

3.2.5 Espacios de datos (Data Spaces)

El término "espacio de datos" abarca varios conceptos relacionados que permiten gestionar todas las fuentes de datos de una organización, independientemente de su formato o ubicación. Estos espacios modelan las relaciones entre las distintas fuentes de datos y son útiles para la gestión de datos en diversos dominios y regiones. Se han propuesto múltiples definiciones, destacando la descentralización y la confianza en el intercambio de datos. Además, han surgido soluciones como *Industrial Data Spaces* (IDS) y *Personal Data Spaces* (PDS) para abordar la regulación y el control de datos privados y personales, en respuesta a la creciente desconfianza en grandes empresas tecnológicas (Curry et al., 2022, 3-7).

Los Espacios de Datos Comunes Europeos están diseñados para aumentar la disponibilidad de datos para su acceso y reutilización en un entorno confiable y seguro, en beneficio de las empresas y ciudadanos europeos. Estos espacios de datos comunes se alinean con la estrategia europea para los datos de febrero de 2020, estableciendo un camino hacia la creación de espacios de datos comunes en diversos sectores estratégicos como salud, agricultura, manufactura, energía, movilidad, finanzas, administración pública y ciencia abierta europea, entre otros. El objetivo es desbloquear el enorme potencial de la innovación impulsada por datos, permitiendo que los datos de toda la UE estén disponibles y se

intercambien de manera confiable y segura. Estos espacios de datos no solo fomentan la innovación, sino que también fortalecen el desarrollo de nuevos productos y servicios basados en datos en la UE, formando el núcleo de una economía de datos europea interconectada y competitiva. En 2024, los interesados continuarán trabajando en el despliegue de estos espacios de datos comunes, respaldados por iniciativas financiadas por la UE y el desarrollo de instrumentos legislativos habilitadores (Comisión Europea, s.f.).

La estrategia europea para los datos busca crear un mercado único para ellos, asegurando la competitividad y soberanía de Europa. Esto incluye permitir que los datos fluyan dentro de la UE y entre sectores, con reglas claras y justas para su acceso y uso. Los Espacios Comunes de Datos Europeos se establecerán para garantizar más disponibilidad de datos en la economía y la sociedad, manteniendo el control sobre ellos. Estos espacios también serán fundamentales para la implementación de técnicas de inteligencia artificial y el apoyo al mercado de servicios en la nube y en el borde, con acciones de investigación e innovación ya financiadas para desarrollar plataformas de datos interoperables (Curry et al., 2022, 7-9).

Los datos son esenciales para el desarrollo de la inteligencia artificial (IA), y la Unión Europea (UE) busca facilitar su acceso mediante la creación de un Espacio de Datos Común Europeo en el sector público. Esto permitirá la reutilización de datos del sector público para impulsar la innovación, el crecimiento económico y la superación de desafíos sociales. La UE propone legislar para abrir más datos para su reutilización y facilitar el intercambio de datos entre el sector público y privado. Se enfatiza la importancia de respetar las normas de protección de datos personales mientras se promueve la innovación. La creación de espacios de datos comunes europeos en áreas clave como manufactura o energía será crucial para el desarrollo de nuevos productos y servicios basados en datos. Para garantizar un acceso, intercambio y reutilización efectivos de estos conjuntos de datos, se requiere el desarrollo y la adopción de reglas de interoperabilidad y estándares (Nikolinakos, 2023, 78-82).

Los "*data markets*" se refieren a entornos donde los datos se compran, venden o intercambian como activos comerciales, con participantes que buscan beneficios económicos mediante transacciones comerciales directas. Por otro lado, los "*data spaces*" se centran en la colaboración y el intercambio de datos entre múltiples partes interesadas, con un enfoque en la interoperabilidad y la colaboración sin necesariamente implicar transacciones comerciales directas. Mientras que los "*data markets*" se centran en la transacción comercial de datos, los "*data spaces*" están orientados hacia la colaboración y el intercambio de datos para diversos fines, no necesariamente comerciales.

El Reglamento de Gobernanza de Datos (2022), destaca la importancia de los espacios de datos en la economía digital, resaltando su potencial para la innovación y el crecimiento económico. Se menciona la creación de un mercado único europeo de datos, donde estos puedan ser utilizados independientemente de su ubicación física, promoviendo así el intercambio de datos y la interoperabilidad. Se hace hincapié en la necesidad de garantizar la seguridad y la protección de los datos, así como en fomentar la participación de diversos actores, incluidas las PYME y los *startups*. Además, se propone la creación de un marco regulatorio armonizado para el intercambio de datos en el mercado interior, así como la promoción de la cooperación entre los Estados miembros. También se menciona el papel clave de los servicios de intermediación de datos para facilitar el intercambio voluntario de datos entre empresas y cumplir con las obligaciones legales. Por último, se propone la creación de una Junta Europea de Innovación de Datos para asesorar a la Comisión en cuestiones relacionadas con la regulación y el desarrollo de los espacios de datos en Europa.

Por su parte, el Reglamento de Datos (2023), expone la importancia de la estandarización y la interoperabilidad semántica para garantizar soluciones técnicas que aseguren la interoperabilidad dentro y entre los espacios de datos europeos comunes. Estos espacios son marcos interoperables específicos por sector o propósito, que comparten estándares y prácticas para el intercambio o procesamiento conjunto de datos, con el fin de desarrollar nuevos productos y servicios, investigación científica o iniciativas de la sociedad civil. La regulación

establece requisitos esenciales para la interoperabilidad, los participantes deben cumplir con estos requisitos en la medida de lo posible. Se sugiere que la conformidad puede lograrse cumpliendo con los estándares armonizados o especificaciones comunes a través de una presunción de conformidad. La Comisión evaluará las barreras a la interoperabilidad y priorizará las necesidades de estandarización, pudiendo solicitar a las organizaciones europeas de estandarización la elaboración de estándares armonizados que satisfagan los requisitos esenciales de la regulación. En caso de que esto no sea suficiente, la Comisión podrá adoptar especificaciones comunes en esas áreas, siempre respetando el papel y funciones de las organizaciones de estandarización. Además, se establece el papel del Consejo de Innovación de Datos Europeos (EDIB) para asesorar y facilitar la aplicación consistente de la regulación, así como su participación en los procesos de estandarización y adopción de especificaciones comunes.

En conclusión, los Espacios de Datos, tanto a nivel europeo como global, representan un cambio fundamental en la forma en que las organizaciones gestionan y comparten datos. Estos espacios no solo abordan la necesidad de interoperabilidad y confianza en el intercambio de datos, sino que también tienen como objetivo desbloquear el potencial de la innovación impulsada por datos. A través de iniciativas como los Espacios Comunes de Datos Europeos, la UE busca promover la competitividad y la soberanía de Europa en el ámbito de los datos, facilitando su acceso y reutilización para impulsar el crecimiento económico y abordar desafíos sociales. Además, al distinguir entre "*data markets*" y "*data spaces*", se reconoce la importancia de la colaboración y la interoperabilidad en la era de la información, no solo para fines comerciales, sino también para promover el progreso y el bienestar general.

Por ejemplo, un experto en espacios de datos puede coordinar un Espacio de Datos Común Europeo en salud, asegurando la interoperabilidad y el acceso seguro a datos anonimizados entre hospitales, laboratorios y empresas farmacéuticas. Cumple con el Reglamento de Gobernanza de Datos y el RGPD,

protegiendo los datos personales de los pacientes, mientras fomenta la innovación médica y garantiza seguridad y confianza entre los participantes.

3.2.6 Ecosistemas de datos (Data Ecosystems)

Un ecosistema de datos es un sistema socio técnico que permite extraer valor de las cadenas de valor de los datos, facilitando la colaboración entre diversas organizaciones e individuos. Con la transformación digital, se está creando un vasto ecosistema de datos que abarca todos los aspectos de nuestra vida, desde datos estructurados hasta imágenes y texto. Este entorno requiere integrar datos de múltiples fuentes, como demuestran las ciudades inteligentes, donde diferentes sistemas colaboran para optimizar las operaciones urbanas. Para aprovechar al máximo estos sistemas, es esencial fomentar el intercambio de conocimientos entre los participantes y repensar la forma en que se comparten los datos en estos entornos complejos (Curry et al., 2022, 2-3).

Los ecosistemas de datos son subelementos de ecosistemas digitales más amplios, que incluyen no solo datos, sino también estándares para la creación e interoperabilidad de conjuntos de datos, aplicaciones, servicios y mecanismos de pago, además de protocolos para interactuar con otros ecosistemas digitales (WEF, 2021f, 11). Incluso, se han propuesto ecosistemas de datos abiertos, que va de la mano de un derecho universal de acceso y reutilización de datos para promover la sostenibilidad (Ploeger & Van Loenen, 2018, 282-285).

Aunque para algunos autores declaran que existe una complejidad de los ecosistemas de datos y sus relaciones constituyentes, la cual hace difícil monitorear manualmente cada aspecto, pues intentar hacerlo impone costos administrativos prohibitivamente altos tanto para la toma de decisiones del sujeto de datos como para el cumplimiento por parte del recolector de datos (WEF, 2021g, 7). Por otra parte, varios países están enfocados en aprovechar los ecosistemas de datos para impulsar la adopción de la IA en el sector público. Colombia prioriza el apoyo a entidades gubernamentales en la adopción de soluciones tecnológicas. Estonia ha desarrollado #KrattAI, una red de aplicaciones

de IA interoperables para que los ciudadanos accedan a servicios públicos mediante asistentes virtuales. Hungría busca dar a los ciudadanos control total sobre sus datos con la introducción de billeteras de datos y servicios personalizados. Los Emiratos Árabes Unidos han establecido un laboratorio de IA para probar casos de uso en diversos servicios urbanos. Además, países como Estados Unidos, Ámsterdam y Helsinki están implementando registros de IA para rastrear cómo se utilizan los algoritmos en las ciudades (OCDE, 2021, 43-44).

La diferencia fundamental entre un ecosistema de datos y un espacio de datos radica en su enfoque y alcance. Un ecosistema de datos es un sistema socio técnico que permite extraer valor de las cadenas de valor de los datos, con la participación de múltiples organizaciones e individuos, tanto a través de competencia en el mercado como de colaboración. Por otro lado, un espacio de datos se refiere a un conjunto de fuentes de datos para una organización, independientemente de su formato, ubicación o modelo, modelando las relaciones entre estas fuentes. Mientras que un ecosistema de datos se enfoca en el valor general de los datos en un contexto más amplio y colaborativo, un espacio de datos se concentra en la gestión y modelado de las relaciones entre fuentes de datos específicas dentro de una organización o sistema (Curry et al., 2022, 2-3).

En resumen, los ecosistemas de datos representan sistemas socio técnicos que permiten la extracción de valor de las cadenas de valor de los datos, promoviendo la colaboración entre diversas entidades. Este concepto se entrelaza con la transformación digital, dando lugar a vastos entornos de datos que abarcan todos los aspectos de la vida moderna. A medida que evolucionan, es crucial fomentar el intercambio de conocimientos y repensar las prácticas de intercambio de datos en estos complejos entornos. Por otro lado, es importante diferenciar entre los ecosistemas de datos y los espacios de datos, siendo los primeros más amplios y colaborativos, mientras que los últimos se centran en la gestión de datos dentro de una organización específica.

Por ejemplo, un profesional en ecosistemas de datos diseña y coordina un ecosistema para una ciudad inteligente, integrando datos de tráfico, transporte, energía y emergencias. Establece estándares de interoperabilidad y un modelo de

gobernanza que garantizan seguridad, eficiencia, transparencia y confianza, optimizando recursos y beneficiando a los ciudadanos.

3.2.7 Cultivos de Datos (Data Cultivars)

En el ámbito público y académico, los datos recopilados de los usuarios individuales de tecnologías de información y comunicación se presentan como flujos brutos de observaciones que son analizados y mejorados por intermediarios de datos y empresas de plataformas. Sin embargo, esta narrativa de "datos crudos" ha generado críticas considerables, ya que la recopilación de datos inevitablemente implica decisiones sobre qué recopilar, qué unidades de medida utilizar y qué formatos y metadatos emplear. Además, los avances en técnicas de minería de datos permiten un análisis más allá de las categorías predefinidas, lo que desafía la idea tradicional de vigilancia como imposición de disciplina artificial. Aun así, los datos recopilados no son simplemente dados, sino que son cultivados y seleccionados activamente para su valor comercial, similar a la agricultura industrial que selecciona variedades de granos adecuadas para la producción a gran escala. Los datos resultantes son tanto crudos como cultivados, tanto reales como altamente artificiales (Cohen, 2019, 64-66).

El término "*data cultivars*" emerge en un contexto donde los datos, lejos de ser vistos como meros flujos brutos de información, son considerados productos que han sido activamente seleccionados y desarrollados para su valor comercial. En el ámbito público y académico, la narrativa predominante sobre los datos tiende a presentarlos como observaciones en bruto que son refinadas por intermediarios y empresas de plataformas. Sin embargo, esta concepción ha sido objeto de críticas significativas, ya que la recopilación de datos implica una serie de decisiones sobre qué recopilar, cómo medirlo y qué formatos utilizar, entre otros aspectos.

La evolución de las técnicas de minería de datos ha ampliado aún más este panorama al permitir un análisis más allá de las categorías predefinidas, desafiando así la noción tradicional de vigilancia como una imposición de

disciplina artificial. Como anteriormente se mencionó, este término sugiere una analogía con la agricultura industrial, donde los datos son tratados como cultivos que son seleccionados y desarrollados activamente para su valor comercial.

Al igual que en la agricultura, donde se seleccionan variedades de granos adecuadas para la producción a gran escala, en el ámbito de los datos, se realizan decisiones activas sobre qué datos recopilar, cómo procesarlos y qué características enfatizar para maximizar su valor. Así, los datos resultantes son tanto crudos como cultivados, representando una combinación de observaciones reales y construcciones altamente artificiales.

Aunque el término *data cultivars* aún no ha sido ampliamente adoptado, su propuesta despierta importantes reflexiones sobre la naturaleza de los datos en la era digital. Reconocer que los datos no son simplemente crudos, sino que son cultivados activamente, puede llevar a una comprensión más profunda de los procesos de recopilación, procesamiento y uso de datos en la sociedad contemporánea.

Por ejemplo, un experto en cultivos de datos (data cultivars) podría trabajar en una empresa de comercio electrónico que busca optimizar sus recomendaciones de productos. En lugar de recopilar datos de comportamiento del usuario de manera generalizada, el experto selecciona activamente qué datos específicos son más valiosos, como historial de compras, tiempo de permanencia en ciertas categorías y clics en promociones. Luego, diseña procesos de recopilación y etiquetado que maximizan la calidad y relevancia de estos datos, integrando metadatos que faciliten su análisis. Con técnicas avanzadas de minería de datos, identifica patrones que no estaban previamente categorizados, permitiendo crear algoritmos de recomendación altamente personalizados. Así, los datos crudos se transforman en un recurso refinado y de alto valor comercial.

3.2.8 Refinerías de Datos (Data Refineries)

Después de ser cultivados y recolectados, los datos personales son procesados para generar patrones y predicciones sobre las preferencias y comportamientos de los sujetos de datos. Estas prácticas de procesamiento de datos, al igual que las de extracción y contratación discutidas previamente, suelen estar envueltas en secreto. Sin embargo, es posible entender el papel de estos procesos dentro de la narrativa imaginada del dominio público biopolítico sin necesidad de acceder a los detalles técnicos. Los refinamientos de datos, en la economía política del capitalismo informativo, funcionan como refinerías de la era de la información, convirtiendo los flujos de datos en formas óptimas para su explotación a gran escala. Estos refinamientos buscan hacer que los comportamientos humanos sean calculables, predecibles y rentables en conjunto, subsumiendo la variación individual dentro de un gradiente radicalmente behaviorista. Estas representaciones virtuales, llamadas "datos dobles", se utilizan para gestionar y comercializar poblaciones, lo que les otorga un carácter biopolítico. Los datos dobles se comercializan en mercados especializados con vida institucional propia, similar a los mercados derivados de productos agrícolas en la industria alimentaria, lo que refleja una innovación extraordinaria pero también oculta la dependencia de la monocultura y dificulta abordar sus efectos externos en la salud humana y ambiental (Cohen, 2019, 66-68).

En la práctica, un refinador de datos es un repositorio capaz de recibir, procesar y transformar datos poliestructurados dispares en formatos utilizables para análisis. Se puede comparar un refinador de datos con una refinería de petróleo, donde se sabe cómo producir gasolina y queroseno a partir del petróleo. Sin embargo, un refinador de datos es más estricto en los datos que acepta para análisis, con un énfasis en evitar que datos irrelevantes contaminen el proceso. Además, puede manejar conjuntos de datos extremadamente grandes de cualquier formato de manera rentable (Trujillo et al., 2015, 20).

La diferencia fundamental entre "*Data Cultivar*" y "*Data Refineries*" radica en las etapas del proceso de manipulación de datos que describen. El "Data Cultivar" se refiere a la recolección inicial y la cruda acumulación de datos de diversas fuentes, similar a sembrar y cosechar en la agricultura, mientras que las "*Data*

Refineries" representan la fase posterior en la que estos datos son procesados, analizados y refinados para generar información más específica y útil. Es como llevar los cultivos agrícolas a una refinería donde se transforman en productos finales útiles. En resumen, el "*Data Cultivar*" se centra en la recolección inicial de datos crudos, mientras que las "*Data Refineries*" se enfocan en el procesamiento y refinamiento de esos datos para obtener información más valiosa y específica.

Finalmente, en un mundo donde los datos personales se han convertido en una materia prima crucial, la analogía entre "*Data Cultivar*" y "*Data Refineries*" ofrece una visión esclarecedora de los procesos involucrados en su manipulación. Mientras que el primero se asemeja a la siembra y cosecha de datos, es en las "*Data Refineries*" donde estos datos crudos son transformados en formas óptimas para su explotación a gran escala. Como en una refinería de petróleo, donde se convierte crudo en productos útiles, las "*Data Refineries*" toman datos poliestructurados dispares y los procesan para generar información específica y valiosa. Esta distinción subraya la importancia de entender las distintas etapas del proceso de manipulación de datos en la economía política del capitalismo informativo, donde la transformación de datos crudos en información refinada impulsa la gestión y comercialización de poblaciones, reflejando tanto innovación como desafíos en la protección de la salud humana y ambiental.

Por ejemplo, un profesional en refinerías de datos (*data refineries*) podría trabajar en una empresa de seguros que recopila grandes volúmenes de datos de dispositivos IoT instalados en vehículos de sus clientes. Después de recibir datos crudos como velocidad, ubicación y frecuencia de frenado, el experto los procesa en una refinería de datos para eliminar ruido e inconsistencias. Luego, utiliza algoritmos avanzados para transformar esos datos en patrones refinados, como índices de conducción segura o perfiles de riesgo. Esta información optimizada se utiliza para personalizar las primas de seguro, incentivar hábitos de conducción seguros y mejorar la rentabilidad del negocio al predecir y mitigar riesgos potenciales con mayor precisión.

3.2.9 Grupos de Datos (Data Pools)

Los *data pools*, en español "agrupamientos o grupos de datos", son una práctica donde empresas combinan sus datos en un único y amplio conjunto, con el objetivo de obtener perspectivas e información que de otra manera no podrían obtener de manera independiente. Este concepto ha sido abordado por la Comisaria de la Unión Europea Margrethe Vestager, quien lo considera una conducta pro competitiva, capaz de aliviar problemas de competencia en grandes plataformas e incluso aumentar la competencia. Sin embargo, también señala la necesidad de utilizar las Directrices Horizontales para referencia al agrupar datos, ya que el intercambio de información podría considerarse anticompetitivo. Es crucial que los data pools sean de acceso abierto para no incurrir en violaciones de las regulaciones de competencia de la UE (Lundqvist, 2023, 89-92).

De manera más general, los data pools son herramientas que permiten la agregación y compartición de datos entre diferentes entidades dentro de una industria o mercado específico. Estas entidades pueden incluir empresas, instituciones gubernamentales o cualquier otra organización que recopile datos relevantes. La idea detrás de los *data pools* es que, al combinar y compartir datos de manera colaborativa, se pueden obtener beneficios, como la creación de nuevos productos o servicios, el desarrollo de soluciones innovadoras, la mejora de la calidad de los productos existentes y la prevención de monopolios al distribuir el poder de los datos entre múltiples actores. Sin embargo, también existen preocupaciones sobre posibles prácticas anticompetitivas y problemas de privacidad que deben abordarse al implementar data pools (Ballardini, 2019, 459).

Cabe destacar que, ante el Reglamento de Gobernanza de Datos (2022), se propone la creación de espacios europeos de datos comunes, conocidos como data pools, para facilitar el intercambio de datos en áreas como la salud, la movilidad, la energía y la agricultura. Estos espacios buscan hacer que los datos sean accesibles, interoperables y reutilizables, promoviendo así la innovación y el crecimiento económico. Además, se destaca la importancia de establecer organizaciones de altruismo de datos reconocidas para fomentar el uso de datos

con objetivos de interés general, como la investigación científica y la lucha contra el cambio climático. Se debe subrayar que para la presente investigación se separaron los *data commons*, *data spaces* y *data pools*, indistintamente de que el reglamento en comento los considere símiles.

En resumen, los data pools representan una estrategia prometedora para fomentar la colaboración y el intercambio de datos entre diversas entidades dentro de una industria o mercado específico. Si se implementan adecuadamente, pueden conducir a la generación de nuevas ideas, productos y servicios, así como a una mayor competencia al distribuir el poder de los datos de manera más equitativa. Sin embargo, para asegurar su efectividad y cumplir con las regulaciones de competencia y privacidad, es fundamental establecer directrices claras y garantizar un acceso abierto y justo a los datos compartidos. De este modo, los data pools pueden desempeñar un papel crucial en el impulso de la innovación y el desarrollo económico, siempre y cuando se aborden cuidadosamente sus posibles implicaciones anticompetitivas y de privacidad.

Por ejemplo, un profesional en grupos de datos (*data pools*) podría trabajar en un consorcio de empresas agrícolas que busca optimizar la gestión de recursos hídricos en zonas rurales. Este experto lidera la creación de un *data pool* donde las empresas comparten datos como niveles de humedad del suelo, patrones de lluvia y consumo de agua. Estos datos combinados permiten desarrollar modelos predictivos para asignar agua de manera más eficiente, reducir costos y minimizar el impacto ambiental. El experto asegura que el *data pool* cumpla con las Directrices Horizontales de la UE, estableciendo políticas claras de acceso abierto y garantizando que no se infrinjan regulaciones de privacidad o competencia.

3.2.10 Uniones de Datos (Data Unions)

Los *Data Unions*, o Uniones de Datos, son estructuras organizativas que permiten a los individuos agrupar y compartir sus datos para su monetización de manera colectiva y justa. Estas uniones, actúan como intermediarios de datos, permitiendo a los usuarios dirigir sus datos hacia grupos específicos que consideren

beneficiosos en términos de privacidad, confianza, ingresos y apoyo a organizaciones o causas con las que estén alineados. Los *Data Unions* operan bajo un marco de gobernanza que garantiza que los beneficios generados por la venta de datos se redistribuyan de manera equitativa entre los miembros, mientras que los operadores de la unión tienen la responsabilidad fiduciaria de actuar en el mejor interés de estos. Además, estas estructuras están diseñadas para evitar la monopolización o el abuso de los datos, asegurando que los intermediarios no puedan ofrecer análisis de datos utilizando la misma información que se vende (Thompson, 2023, 115-116).

En el mismo sentido, estas son propuestas que reconocen los límites de una visión individualista de la protección de datos y la privacidad. Se basan en un enfoque más colectivo hacia la gobernanza de datos, tratando los datos como un recurso compartido en lugar de individual. Al igual que un sindicato laboral, un *data union* actuaría como una herramienta de negociación colectiva para las personas que generan datos, negociando mejores términos y, posiblemente, compensaciones económicas con grandes empresas tecnológicas. Aunque el concepto aún está en evolución, se considera una forma de empoderar a los individuos en la economía de datos. Sin embargo, surgen desafíos sobre cómo financiarlos sin comprometer su independencia y determinar sus propósitos permitidos. Por ejemplo, si se debiera permitir la formación de *data unions* con el propósito principal de maximizar beneficios o si deben tener un enfoque más altruista hacia la protección de los datos de sus miembros (Renieris, 2023, 73-74).

Los *data pools* implican la agregación y compartición de datos entre múltiples actores dentro de una industria o mercado específico, con el objetivo de obtener beneficios mutuos como el desarrollo de productos o análisis de mercado. Por otro lado, las *data unions* se centran en representar los intereses de los individuos en la gestión y monetización de sus datos personales, actuando como intermediarios entre los individuos y las empresas para garantizar sus derechos y recibir una compensación justa por el uso de sus datos. Mientras los data pools se enfocan en la colaboración empresarial, las *data unions* priorizan la protección de los derechos individuales sobre los datos.

En conclusión, los *Data Unions* representan un paradigma emergente en la gestión de datos que desafía la noción tradicional de propiedad individual en favor de un enfoque más colectivo y justo. Al actuar como intermediarios entre individuos y empresas, estos modelos organizativos buscan garantizar la equidad en la monetización de datos personales, mientras protegen los derechos individuales. A diferencia de los Data Pools, que se centran en la colaboración empresarial, las *Data Unions* ponen énfasis en la protección y empoderamiento de los usuarios en la economía de datos, promoviendo una visión más equitativa y democrática de la gestión de la información. Sin embargo, persisten desafíos en torno a su financiamiento y propósito, lo que sugiere la necesidad de un debate continuo sobre su implementación y regulación.

Por ejemplo, un experto en uniones de datos puede liderar la creación de una unión para usuarios de dispositivos de salud portátiles, como relojes inteligentes. Reúne a usuarios interesados en compartir datos de salud, como ritmo cardíaco y patrones de sueño, con empresas de investigación médica. Negocia contratos para garantizar la anonimización, el uso ético y la monetización justa de los datos, redistribuyendo los ingresos entre los miembros. Los usuarios pueden elegir qué investigaciones apoyar, como estudios de enfermedades crónicas. Además, establece un marco de gobernanza transparente para proteger los derechos de los usuarios y fomentar su confianza en la unión.

3.2.11 Datos comunes y clubes de datos (Data commons and Data Clubes)

Un "*data commons*" es una red de relaciones entre titulares de derechos de datos que tienen igual derecho a un recurso de datos común e indivisible. Esta estructura sigue los ocho principios de Elinor Ostrom para gobernar los bienes comunes, que incluyen la claridad en las fronteras de usuarios y recursos, la congruencia entre beneficios y costos, la capacidad de los usuarios para establecer sus propias reglas, la vigilancia regular de usuarios y condiciones del recurso, la aplicación de sanciones graduales, mecanismos de resolución de conflictos, un mínimo reconocimiento de derechos por parte del gobierno y

empresas anidadas. Los datos son un recurso indivisible y los miembros tienen derechos iguales sobre ellos, lo que significa que los datos permanecen inalterados a pesar de que los miembros se unan o abandonen el grupo (WEF, 2022, 15).

Los datos comunes son espacios que agregan datos de una amplia gama de fuentes en una base de datos unificada para que sea más accesible y útil. Este es un término que se ha utilizado en los campos de la investigación y atención médica. Los datos comunes han proliferado en todo el mundo y están proporcionando valor a la atención médica e investigación (Thompson, 2023, 116-117).

La idea de los bienes comunes de datos surge como una alternativa al enfoque del mercado libre de datos, proponiendo que los datos se compartan a través de infraestructuras compartidas para hacerlos menos excluyentes para los intereses económicos. Este enfoque, aunque poderoso en la política tecnológica y la gobernanza de datos, aún no se ha materializado completamente. Se propone un modelo de "semi-bienes comunes" donde diferentes entidades tienen derechos condicionales sobre los datos, priorizando la integridad distributiva sobre los intereses de los actores del mercado más poderosos. En la UE, se están desarrollando instrumentos legislativos que reflejan esta filosofía, especialmente en la compartición de datos entre empresas, gobierno y sociedad en general. Sin embargo, la implementación de los bienes comunes de datos requiere adaptaciones debido a las complejidades tecnológicas y arquitectónicas, y debe considerar cuidadosamente las circunstancias específicas para asegurar beneficios para los usuarios previstos (Solano et al., 2022, 27-28).

Si bien en *lato sensu, data commons* y data clubes son lo mismo, en *stricto sensu*, las diferencias entre *Data Commons y Data Clubs* radican en su estructura, objetivos y enfoque respecto a la gestión y compartición de datos. Los *Data Commons* representan un modelo de colaboración abierta donde los datos se comparten libremente para fomentar la innovación y el bien común. Por otro lado, los Data Clubs son grupos más selectivos y cerrados, formados por individuos u organizaciones con intereses comunes en datos específicos, y suelen enfocarse

en objetivos más concretos, como el análisis de datos para obtener *insights* comerciales. Mientras que los *Data Commons* promueven el acceso abierto y la colaboración comunitaria, los Data Clubs tienden a gestionar y compartir datos de manera más restringida entre sus miembros.

En resumen, la idea de los "*data commons*" representa un paradigma emergente en la gestión y compartición de datos, basado en los principios de Elinor Ostrom para gobernar los bienes comunes. A través de una estructura que garantiza derechos equitativos sobre los datos, se busca fomentar la colaboración abierta y el beneficio común. Sin embargo, mientras que los "*data commons*" abogan por la apertura y la colaboración comunitaria, los "data clubs" adoptan un enfoque más selectivo y cerrado, centrado en intereses específicos y objetivos concretos. Esta distinción entre ambos modelos refleja la diversidad de enfoques y estructuras dentro del ámbito de la gestión de datos en la era digital.

Por ejemplo, un experto en datos comunes y clubes de datos gestiona proyectos de colaboración, como un data commons para investigar enfermedades raras, centralizando datos anonimizados de pacientes y garantizando acceso equitativo para la investigación. Define reglas de gobernanza para promover la transparencia, resolver conflictos y asegurar la sostenibilidad. Simultáneamente, organiza un data club para instituciones interesadas en desarrollar tratamientos específicos, restringiendo el acceso a los miembros. Este enfoque combina investigación abierta y colaboración estratégica, maximizando el valor de los datos y fomentando innovación e investigación focalizada.

3.2.12 Cooperativas de Datos (Data Cooperatives)

Una cooperativa de datos es una red de acuerdos entre pares con intereses mutuos que permite la agrupación de recursos de datos. Los miembros aportan datos y son responsables de gestionarlos. Los datos se añaden y eliminan conforme los miembros se unen y abandonan la cooperativa (WEF, 2022, 15).

Las cooperativas de datos son modelos alternativos de gobernanza en los que los sujetos de datos organizan colectivamente el uso, acceso y compartición

de datos. Estas cooperativas se centran en el interés público y la protección de la comunidad, desafiando la idea de los datos como una mercancía de mercado. Buscan una gestión descentralizada y transparente de los datos, y su objetivo es abordar los desequilibrios de poder en la gobernanza de datos. Además, las cooperativas de datos tienen diversos intereses, desde la privacidad y el trabajo hasta preocupaciones sociales como la sostenibilidad. Ejemplos incluyen iniciativas como FairBNB para compartir beneficios entre anfitriones y comunidades locales, y cooperativas de plataformas para trabajadores independientes en servicios de entrega y transporte (Solano et al., 2022, 28-29).

Las cooperativas de datos funcionan como guardianes de los datos de sus usuarios, negociando el acceso justo a estos datos para investigación académica, farmacéutica y otros servicios. Garantizan la publicación de resultados de investigaciones, la devolución de datos a los usuarios y la fijación de precios justos para el acceso a terceros. Este modelo se proyecta como rentable, ya que combina diferentes tipos de datos bajo el control de los individuos, generando un valor de mercado significativo. Los ingresos se reinvierten en el mantenimiento y desarrollo de la plataforma, y los beneficios financieros se destinan a la sociedad en general, en lugar de distribuirse entre los individuos, argumentando que el valor real reside en la contribución colectiva de los datos (Hafen, 2019, 145-146).

Los *Data Unions* y las *Data Cooperatives* son estructuras que permiten la gestión colectiva de datos para beneficio mutuo, pero difieren en su enfoque y estructura. Los *Data Unions* se centran en la monetización de datos, utilizando tecnologías descentralizadas para garantizar una distribución equitativa de los ingresos generados por la venta de datos, mientras que las *Data Cooperatives* pueden tener un alcance más amplio, incluyendo la creación de valor compartido y la colaboración en diferentes áreas. Mientras los *Data Unions* tienden a enfocarse en la monetización equitativa de los datos, las *Data Cooperatives* pueden tener una orientación más diversa hacia la colaboración y el intercambio de conocimientos entre sus miembros.

El Reglamento de Gobernanza de Datos (2022), establece que las cooperativas de datos buscan lograr varios objetivos, especialmente fortalecer la

posición de los individuos para tomar decisiones informadas antes de consentir al uso de datos, influir en los términos y condiciones de las organizaciones de usuarios de datos en relación con su uso de una manera que brinde mejores opciones a los miembros individuales del grupo o posiblemente encontrar soluciones a posiciones conflictivas de los miembros individuales de un grupo sobre cómo se pueden utilizar los datos cuando estos datos se relacionan con varios sujetos de datos dentro de ese grupo. En este contexto, es importante reconocer que los derechos bajo el RGDP (2016) son derechos personales del interesado en los datos y que los interesados en los datos no pueden renunciar a tales derechos.

Asimismo, el Reglamento de Gobernanza de Datos (2022), reconoce que las cooperativas de datos también podrían proporcionar un medio útil para empresas unipersonales y PYMEs que, en términos de conocimiento sobre intercambio de datos, a menudo son comparables a individuos. Los "servicios de cooperativas de datos" se refieren a los servicios de intermediación de datos ofrecidos por una estructura organizativa constituida por sujetos de datos, empresas unipersonales o PYMEs que son miembros de esa estructura, teniendo como principales objetivos apoyar a sus miembros en el ejercicio de sus derechos con respecto a ciertos datos, incluido el hecho de tomar decisiones informadas antes de dar su consentimiento al procesamiento de datos, intercambiar opiniones sobre los propósitos y condiciones de procesamiento de datos que mejor representen los intereses de sus miembros en relación con sus datos, y negociar términos y condiciones para el procesamiento de datos en nombre de sus miembros antes de dar permiso para el procesamiento de datos no personales o antes de dar su consentimiento para el procesamiento de datos personales.

En conclusión, las cooperativas de datos representan un modelo alternativo de gobernanza que desafía la noción tradicional de los datos como una mercancía de mercado. Estas estructuras, como los *Data Unions* y las *Data Cooperatives*, permiten la gestión colectiva de datos para beneficio mutuo, fomentando la transparencia, la equidad y la colaboración entre sus miembros. Al poner énfasis en el interés público y la protección de la comunidad, estas cooperativas buscan

abordar los desequilibrios de poder en la gobernanza de datos y promover un enfoque más descentralizado y transparente en la gestión de la información. Además, al reinvertir los ingresos en el desarrollo de la plataforma y destinar los beneficios financieros a la sociedad en general, las cooperativas de datos destacan el valor de la contribución colectiva de los datos en lugar de centrarse exclusivamente en su monetización.

Por ejemplo, un profesional en cooperativas de datos lidera iniciativas como una cooperativa para agricultores que comparten datos sobre cultivos, uso de agua y fertilizantes. Los miembros contribuyen voluntariamente con sus datos y mantienen control sobre su uso. La cooperativa negocia con empresas agrícolas para intercambiar acceso a los datos por beneficios, como descuentos en insumos o herramientas personalizadas. Además, implementa un marco transparente y descentralizado para decisiones colectivas sobre el uso de los datos, respetando la privacidad y maximizando los beneficios, promoviendo una colaboración justa y sostenible.

3.2.13 Colaboraciones de Datos (Data collaboratives)

Las colaboraciones de datos representan un modelo alternativo frente al enfoque predominante basado en el mercado para la gobernanza de datos. En estas colaboraciones, datos privados de empresas se combinan con datos públicos mediante un tercero independiente, restringiendo el acceso y uso a los miembros de la colaboración. Se utilizan en proyectos que benefician el desarrollo de políticas públicas y empoderan a los participantes al utilizar datos previamente inaccesibles. Aunque siguen principios del modelo dominante, están más desarrolladas y tienen ejemplos concretos en sectores como salud, transporte y humanitario, como muestra la base de datos del NYU GovLab con más de 200 casos (Solano et al., 2022, 26-27).

Los "*data cooperatives*" son una forma emergente de asociación público-privada que permite la colaboración y el intercambio de información entre sectores y actores diversos. Este modelo ha sido utilizado en diversos sectores y

geografías, desde compartir datos sobre enfermedades para acelerar tratamientos hasta aprovechar datos privados de transporte público para mejorar la planificación urbana. Su potencial radica en superar la fragmentación de la información y los problemas de compatibilidad que actualmente limitan las respuestas políticas en áreas como la migración. En resumen, estos mecanismos podrían facilitar el análisis de la situación relacionada con los movimientos de migrantes y refugiados, mejorando así la capacidad de respuesta institucional ante los cambios en los patrones migratorios (Verhulst & Young, 2019, 470-471).

Un *data collaborative* es una relación de intercambio de datos que puede adoptar diversas formas, como interfaces públicas o intermediarios de confianza. Facilita que entidades compartan datos, modelos y experiencia, fomentando el uso público de datos previamente aislados. Principalmente implica entidades privadas compartiendo datos con el sector público o grupos de interés. Los incentivos para las corporaciones incluyen acceso recíproco a datos, perspicacias de investigación, reputación, ingresos y cumplimiento normativo, además de iniciativas de responsabilidad social corporativa y ambiental (WEF, 2022, 15).

En resumen, las colaboraciones de datos, representadas por modelos como los "*data cooperatives*" y los "*data collaboratives*", ofrecen un enfoque alternativo a la gobernanza de datos, promoviendo la colaboración entre entidades públicas y privadas para el beneficio común. Estos modelos, que restringen el acceso y uso de datos a miembros específicos, han demostrado su eficacia en diversos sectores como la salud, el transporte y la planificación urbana. Al superar la fragmentación de la información y fomentar el intercambio de datos previamente inaccesibles, estos mecanismos tienen el potencial de mejorar la capacidad de respuesta institucional en áreas críticas como la migración y el desarrollo de políticas públicas. Además, ofrecen incentivos tanto para entidades privadas como para el sector público, incluyendo acceso a datos recíproco, perspicacias de investigación y beneficios reputacionales, entre otros. En conjunto, estas colaboraciones representan una herramienta poderosa para aprovechar el valor de los datos en beneficio de la sociedad en su conjunto.

Por ejemplo, un experto en colaboraciones de datos lidera proyectos como el de una empresa de telecomunicaciones que comparte datos anonimizados sobre movilidad con una agencia gubernamental de transporte y un instituto de investigación urbana. El objetivo es optimizar rutas de transporte público en ciudades congestionadas. Este profesional colabora con un tercero independiente que gestiona los datos y define reglas claras para garantizar la privacidad y el uso ético. Al combinar los datos, se identifican áreas de alta demanda no atendidas, permitiendo planificar rutas más eficientes, beneficiando a ciudadanos y participantes mientras se fortalecen las políticas públicas de movilidad.

3.2.14 Fideicomisos de Datos (Data Trust)

Una solución legal y no tecnológica para resolver el intercambio de datos es la creación de los fideicomisos de datos, en el cual se argumenta la relevancia de conceptualizar los datos como propiedad para los fines de la ley de fideicomisos, entonces los propietarios de conjuntos de datos podrían transferir la propiedad de ellos a un fideicomiso legal, y la escritura del fideicomiso establecerá los objetivos del este, de acuerdo con los datos que podrán ser compartidos y quienes podrán administrar el fideicomiso con carácter de fiduciario. No obstante, algunos institutos como el *Open Data Institute* han sugerido la transmutación de la figura de los fideicomisos y centrarse más en derechos de propiedad sobre los datos, utilizando el derecho de sociedades y contratos para construir una estructura legal que se centre en usos permitidos de datos compartidos y equilibrar el interés de todas las partes con los objetivos de la confianza de datos (Reed, 2021, 199-201).

El modelo de fideicomiso de datos públicos (*Public data trusts*) implica que una institución pública administre los datos, promoviendo la confianza y el beneficio público sobre la explotación comercial. Este enfoque, ejemplificado por iniciativas como el Plan Digital de Barcelona y el proyecto DECODE en Ámsterdam y Barcelona, busca utilizar los datos para la toma de decisiones democráticas, la innovación social y la mejora de los servicios públicos. Incluye mecanismos de consulta, rendición de cuentas y acuerdos sólidos para compartir

datos con terceros, manteniendo el interés público como prioridad (Solano et al., 2022, 26).

Un fideicomiso de datos es un conjunto de acuerdos basados en la confianza o el derecho contractual que permite a los titulares de datos delegar su control a un fiduciario. Este fiduciario actúa en interés de los beneficiarios, siguiendo deberes de lealtad y cuidado. Puede ser independiente y no necesariamente con fines de lucro. Los datos se gestionan de acuerdo con los intereses de los titulares, y el nivel de participación de los beneficiarios puede variar. El contrato del fideicomiso establece los términos bajo los cuales se toman decisiones sobre el uso y suministro de datos en beneficio de los depositantes y otras partes interesadas (WEF, 2022, 14).

En conclusión, la implementación de fideicomisos de datos emerge como una solución legal y no tecnológica para abordar los desafíos del intercambio de datos. Al conceptualizar los datos como propiedad y delegar su control a un fiduciario, se establece un marco jurídico que equilibra los intereses de los propietarios de datos, los beneficiarios y otras partes interesadas. Mientras que algunos abogan por enfoques más centrados en los derechos de propiedad sobre los datos, los fideicomisos de datos públicos destacan la importancia de la gestión transparente y democrática de los datos, priorizando el interés público sobre la explotación comercial. Este modelo promueve la confianza, la rendición de cuentas y el beneficio colectivo, ejemplificado en iniciativas como el Plan Digital de Barcelona y el proyecto DECODE, que utilizan los datos para fortalecer la toma de decisiones democráticas y mejorar los servicios públicos. En última instancia, los fideicomisos de datos ofrecen un mecanismo flexible y robusto para gestionar y compartir datos en beneficio de la sociedad en su conjunto.

Por ejemplo, un experto en fideicomisos de datos colabora con hospitales para crear un fideicomiso que administre datos de salud anonimizados destinados a la investigación médica. Los hospitales transfieren la propiedad de sus datos al fideicomiso, gestionado por un fiduciario independiente que protege los intereses de los pacientes y asegura el cumplimiento de las normativas de privacidad. El contrato estipula que los datos solo se compartan con investigadores aprobados

para desarrollar tratamientos específicos, prohibiendo su uso comercial. Este modelo fortalece la confianza, garantiza la transparencia y prioriza el beneficio público en la investigación médica.

3.3 Tipos de Interacciones con los datos

En este subcapítulo, se abordarán los distintos tipos de interacciones con los datos, explorando las diversas formas en que los usuarios pueden interactuar con la información. Previamente, se han examinado los enfoques para mejorar el acceso a los datos y las estrategias de intercambio de datos. Por lo tanto, ahora se realizará un análisis más detallado de cómo los individuos y sistemas pueden interactuar con los datos para extraer valor, tomar decisiones informadas y generar conocimiento.

3.3.1 Para alimentar sistemas

Los beneficios y costos de los datos surgen en gran parte de su uso en la alimentación de algoritmos de IA, para proporcionar información y predecir comportamientos. Existe una relación bidireccional entre la IA y los datos: sin datos, la contribución del campo de la IA se limitaría a los sistemas basados en el conocimiento gobernados por "reglas si-entonces"; y sin IA, el valor que podría extraerse de los datos se limitaría a la experiencia humana y la comprensión teórica de los fenómenos del mundo real, solo mejorada con capacidades de cálculo más rápidas y precisas que las máquinas podrían ofrecer. Se pueden derivar enormes beneficios de la IA y el control de los datos, que brindan no solo ganancias económicas, sino también un enorme poder y capacidad para controlar y dar forma al futuro de la tecnología, la economía y la sociedad (UNCTAD, 2021, 41).

La estrategia para los datos y el libro blanco sobre la IA son los primeros pilares de la nueva estrategia digital de la Comisión Europea, aunado a que estos

giran de acuerdo con una visión de antropocentrismo, donde las personas se encuentran en primer lugar de todos los desarrollos tecnológicos. Desde el ámbito legal y vinculante, se cuenta en la UE con el Reglamento de Gobernanza de Datos (2022) y con la Ley de Inteligencia Artificial (2024). Lo que demuestra la pertinencia de marcos jurídicos sólidos y armoniosos entre datos e IA.

Como se mencionó en la introducción de este proyecto de investigación, la estrategia europea de datos tiene como objetivo la creación de un mercado único de datos que garantice la competitividad global y la soberanía de los datos de Europa. Los espacios de datos europeos comunes garantizarán la disponibilidad para su uso en la economía y la sociedad, y al mismo tiempo existirá un ecosistema armonioso entre gobierno, empresas y las personas que generan los datos. Mientras que el libro blanco sobre la IA puntualiza el apoyo de un enfoque regulatorio y orientado a la inversión para promover la adopción de la IA y abordar los riesgos asociados con ciertos usos de esta tecnología, donde el propósito principal es establecer opciones políticas de cómo lograr estos objetivos, dejando fuera su uso militar. Concluye con un exhorto a los Estados miembros, instituciones europeas y toda parte interesada a reaccionar a las opciones del escrito y contribuir a la futura toma de decisiones en este ámbito.

Aunque los desarrolladores e implementadores de la IA ya están sujetos a la legislación en materia de derechos fundamentales como la protección de datos, privacidad, no discriminación, protección de los consumidores, normas sobre la seguridad de los productos y responsabilidad civil, e incluso sea necesario analizar si la legislación actual puede hacer frente a los riesgos de la IA o de lo contrario sea pertinente la creación de una nueva legislación (Comisión Europea, 2020b). Debe quedar claro, que la disponibilidad de datos es fundamental para entrenar a los sistemas de IA, dado que los productos y servicios están evolucionando rápidamente desde el reconocimiento de patrones y la generación de información hasta técnicas de predicción más sofisticadas y, por tanto, mejores decisiones (Comisión Europea, 2020a).

Dicho esto, es crucial tener datos que permitan a los algoritmos basarse en evidencia para el desarrollo futuro. Actualmente no existen estándares uniformes

en términos de acceso, intercambio y protección de datos, creando desafíos particulares para la IA. De facto, gran parte de los datos que ingresan en los algoritmos son de naturaleza patentada y no se comparten de manera muy amplia con la investigación, lo que limita la innovación. La IA requiere grandes conjuntos de datos para probar y mejorar sus capacidades de aprendizaje, y sin acceso a estos será difícil obtener los beneficios de la IA. Es decir, los datos son el combustible que impulsa el motor de la IA y abrir el acceso a los datos ayudará a obtener información que transformará todos los sectores (West & Allen, 2020, 213-214).

En cambio, es necesario detenerse a resolver ciertos asuntos *a priori*. Desde una visión de protección de datos existen diversos debates e interrogantes respecto a los alcances de las restricciones sobre el procesamiento totalmente automatizado y la elaboración de perfiles; cómo respetar el requisito de transparencia y responsabilidad dados los límites técnicos actuales sobre la explicación de algunos procesos de IA como el aprendizaje profundo y las redes neuronales; o cómo se puede lograr prácticamente la limitación, minimización y anonimización del propósito dada la escala de muchas aplicaciones de IA, y de su uso para encontrar patrones previamente no reconocidos en los datos y la sofisticación de las técnicas de análisis de datos masivos. También el cómo cumplir con el requisito de precisión en los datos cuando los procesos de IA sean intrínsecamente probabilísticos y por último, como respaldar el consentimiento significativo para el procesamiento de la IA (Access Now, 2018, 6).

Además, en razón a que la IA tiene diversas aplicaciones en todos los sectores, su ejecución puede utilizarse en vigilancia, identificación, elaboración de perfiles, empujones (*nudges*) y toma de decisiones individuales automatizadas o combinación de todas estas técnicas (Muller, 2020, 26-41), dando como resultado discriminación intencional, sesgo estadístico e inferencias defectuosas o demasiado precisas (Barocas, 2014, 1-4). Puede haber tratamientos que incluyen componentes de IA que manejan datos de personas físicas, como en un modelo de perfilado de *marketing* o electoral, o puede haber tratamientos en los que no aparezcan datos de carácter personal, como podría suceder en un modelo de

predicción meteorológico que recoge datos de estaciones geográficamente distribuidas. Por lo que se debe prestar especial atención a la legitimación para el tratamiento, la información y transparencia, el ejercicio de derechos, las decisiones automatizadas, la exactitud, la minimización de datos, el análisis de la proporcionalidad del tratamiento y las evaluaciones de impacto para la protección de datos y privacidad, así como nuevos instrumentos como la Evaluación de Impacto Algorítmico para medir parámetros de calidad y adecuación (AEPD, 2020a, 2-37).

Dicho lo anterior, por encima de proponer principios que deberán seguirse en la intersección entre datos e IA, una propuesta más rigurosa es asumir un control equilibrado en mando y control entre *hard law* y *soft law*, es decir, que por encima de debatir entre seguir derecho vinculante o indicativo, se exploren nuevas propuestas para regular y desarrollar soluciones materiales entre la IA y los datos.

Verbigracia, es imperante analizar novísimos mecanismos alternativos a la regulación tradicional como la certificación, auditorías, transparencia, entornos de prueba regulatorios, seguros y evaluaciones de impacto algorítmicas, cuya implementación en la IA que utilicen datos será benefactora para diversos bienes jurídicos protegidos.

Primero, se podría analizar como la certificación podría ser una forma de regulación de riesgos y garantía de calidad de productos y servicios, en el cumplimiento de criterios establecidos por asociaciones profesionales, organizaciones de estándares o agencias gubernamentales (Dignum, 2019, 94-98). En el caso de la IA, la certificación acoge dos caminos; uno es la certificación del algoritmo como un objeto de software que especifica sus atributos de diseño o el proceso, así como la experiencia involucrada por los requisitos relacionados con los resultados que fueron monitoreados y evaluados; y por otro lado, es la certificación de la persona o proceso que utiliza el sistema, es decir, el controlador del sistema que toma decisiones, que acreditará que los algoritmos cumplan con el contexto de uso y los criterios específicos (Koene et al., 2019, 43).

Segundo, se podría apostar a crear escenarios donde las auditorias de inspección adopten un estudio de auditoría tradicional para aplicarla en los

sistemas algorítmicos, dirigida de una hipótesis particular sobre un sistema al observar sus entradas y salidas (inputs y outputs), por ejemplo, para observar si un sistema genera prejuicios raciales en los resultados de una decisión (auditoría de sesgo); o por otro lado, buscar una auditoría de inspección y cumplimiento (como una auditoría financiera), utilizada para describir una inspección integral y verificar si un sistema se comporta de acuerdo con reglas o normas (auditoría de inspección regulatoria con un enfoque amplio) (Ada Lovelace Institute, 2020, 8).

Tercero, podría adoptarse por la transparencia en los sistemas de IA y los datos desde un enfoque técnico para rendir cuentas en: 1. en los datos utilizados por el sistema, ya sean fuentes, procesos y métodos de verificación para valorar imparcialidad y representación; 2. algoritmos, a fin de verificar los resultados de salida con los de entrada e inspeccionar errores internos y externos; 3. objetivos, con el propósito de conocer las prioridades relativas de cada sistema; 4. resultados, para analizar los efectos del despliegue de los sistemas algorítmicos, incluidos los estados internos del sistema; 5. cumplimiento, para determinar el desempeño general de los requisitos de transparencia; 6. influencia, para averiguar si algún elemento del proceso algorítmico se inclinó a favorecer un resultado; y 7. uso, comparando qué datos personales está usando el sistema, ya sea para personalizar los resultados o datos que pueden entrenar al sistema y no se violen los principios de protección de datos y privacidad (Koene et al., 2019, 5-6).

Cuarto, otra visión propositiva para funcionamiento fiable entre los datos y la IA, es considerar los entornos de prueba regulatorios, o *sandboxes*, consintiendo en que nuevas ideas, productos y servicios se puedan probar en un entorno activo, pues las entidades reciben la autorización para experimentar con sus productos o estrategias sin estar sujetas a requisitos existentes, ni a las prohibiciones asociadas o costos de cumplimiento (Fenwick *et al.*, 2018, 82-90). O por su parte, también considerar como un medio de control de riesgo de transferencia, a los seguros, principalmente para determinar con qué tipo de régimen de responsabilidad está vinculada la IA, así como identificar los sujetos responsables relacionados con diferentes roles en su desarrollo, procesamiento,

propiedad, aplicación y ejecución; esto cuestiona si las normas tradicionales serán suficientes para generar responsabilidad legal ante daños ocasionados por sistemas de IA, y menoscaba los principios relativos a la indemnización y reparación del daño causado (Maqueo *et al.*, 2021, 78).

A *prima facie*, pudiera considerarse que un capítulo dedicado al uso de IA no cuenta con armonía con el tema principal de tesis, a pesar de que los datos y los algoritmos son elementos complementarios. Sin embargo, se coincide con Steinrötter (2020, 295) en que probablemente los datos ni siquiera sean el punto de partida adecuado para regular la economía de datos, tal vez se deba mirar a la divulgación de los métodos y técnicas utilizados por los algoritmos y esta sea la solución. Además, la proposición científica radica en explorar nuevos instrumentos de regulación, control y mando sobre la IA y los datos.

3.3.2 Para transmitir y transferir datos

Sólo bajo la guía combinada de tecnología digital, por ejemplo, *Blockchain* y las buenas leyes y gobernanza representada por los derechos de los datos, se puede avanzar en la dirección para mejorar el acceso e intercambio de datos (Yuming, 2021, 7). Aunque esto debería abordarse con detenimiento, ya que se puede utilizar la tecnología para cambiar la naturaleza no rival de los datos y hacerlos excluibles, asegurando que no puedan ser utilizados a menos de que el titular de estos así lo considere (Reed, 2021, 178).

A pesar de que la teoría tradicional de propiedad no sea suficiente para los datos, un posible hito como enfoque de mejora de acceso es por medio de la tecnología *Blockchain* (Treder, 2019, 31), ya que demuestra ser una innovación que podría superar diferentes problemas legales con respecto a la asignación y control de los datos digitales. En primer lugar, mediante *Blockchain* se pueden especificar claramente los activos sobre los que se debe ejercer control, satisfaciendo así el principio de especificidad. Además, la referencia para asignación de datos digitales estaría de acuerdo con el principio de publicidad, pues no solo permiten la verificación del derecho genuino sobre los datos, sino

que también podrían servir como una especie de registro público, haciendo que las transferencias posteriores de la propiedad de los datos a otra parte sean reconocibles para terceros. Por lo tanto, *Blockchain* podría servir como base tecnológica para el comercio de datos, sin limitarse a la esfera *inter vivos*, sino que también podría aplicarse al patrimonio digital de una persona física y servir como un complemento valioso a la iniciativa de autorregulación existente (Chrobak, 2018, 269).

Se debe hacer hincapié en la importancia creciente de la tecnología *Blockchain* en el seguimiento y seguridad de los datos en su movimiento a través de diversas etapas. Esta tiene una relevante trazabilidad de los datos y su procedencia, así como la necesidad de comprender su uso y quién lo está utilizando. Además, la naturaleza de *Blockchain* y sus elementos clave, habilita la confianza en los datos, ofreciendo un registro permanente y distribuido de transacciones. Por último, es importante mencionar la característica de seguridad que proporciona *Blockchain* (Mattsson, 2022, 183-190).

Incluso nace un nuevo servicio basado en computación en la nube y *Blockchain*, llamado *Blockchain-as-a-Service (BaaS)*, el cual es una extensión de *Software as a Service* (SaaS) en la nube, donde una plataforma *Blockchain* se implementa para una organización. El proveedor de la nube gestiona el software y la infraestructura del *Blockchain*, permitiendo que la empresa se centre en sus aplicaciones de negocio sin preocuparse por otros aspectos de la infraestructura. BaaS se puede considerar como un SaaS donde el software es un *Blockchain*, lo que significa que todos los servicios están gestionados externamente. Además, los clientes obtienen una red *Blockchain* completamente gestionada sobre la cual pueden construir y gestionar sus propias DApps. Varias empresas ofrecen servicios BaaS, como AWS, Azure, Oracle e IBM (Bashir, 2020, 678).

Por otra parte, desde un enfoque legal, el Reglamento de Datos (2023) contempla los *smart contracts* o contratos inteligentes, estableciendo normas armonizadas sobre la disponibilidad de datos, incluyendo compensaciones razonables que pueden incluir costos técnicos como reproducción, almacenamiento y procesamiento de datos para facilitar el intercambio de datos.

Estas compensaciones pueden variar según el volumen de los datos y pueden reducirse en transacciones regulares o repetitivas mediante acuerdos a largo plazo, como suscripciones o el uso de *smart contracts*. Para promover la interoperabilidad de herramientas para la ejecución automatizada de acuerdos de intercambio de datos, se establecen requisitos esenciales para los *smart contracts*, como robustez, terminación segura, archivo de datos y continuidad, control de acceso y consistencia con los términos del acuerdo de intercambio de datos. Los desarrolladores profesionales de *smart contracts* deben cumplir con estos requisitos esenciales, y se prevé una presunción de conformidad para los *smart contracts* que cumplan con estándares armonizados. Además, se establecen especificaciones comunes para garantizar la interoperabilidad de los servicios de procesamiento de datos y de los propios *smart contracts*, y se confieren poderes de ejecución a la Comisión para adoptar especificaciones comunes y publicar referencias de estándares armonizados.

En resumen, las grandes intermediarias de internet, que a menudo son proveedoras de servicios en la nube, han logrado convertirse en nodos de transferencia de datos en sus ecosistemas digitales, controlando y privatizando grandes cantidades de información a través de contratos que dificultan la competencia y el funcionamiento de los mercados. Por lo tanto, se necesitan leyes de competencia adecuadas, regulaciones sectoriales y legislación sobre propiedad intelectual para proteger los mercados y la economía liberal en la era de la economía impulsada por los datos (Lundqvist, 2023, 51-56).

En este sentido la UE, la Comisión Europea (2019) ha adoptado una nueva Estrategia en la Nube (2019). La estrategia describe cómo la computación en la nube está dando forma al futuro de las TIC y es un facilitador para la estrategia digital general de la Comisión Europea. En el centro de la estrategia en la nube se encuentra un enfoque de "nube primero" con una oferta de servicios de nube híbrida y multicloud segura. Se resaltan los impactos en las soluciones digitales, el ecosistema de datos, el lugar de trabajo digital, la infraestructura de TI y, especialmente, los servicios de seguridad de TI. La estrategia tiene como objetivo ofrecer servicios en la nube para la Comisión Europea que sean seguros, híbridos,

multicloud y energéticamente eficientes, cumpliendo con las regulaciones de protección de datos y los objetivos de sostenibilidad de la UE.

Conclusiones

Conclusiones

En conclusión, se abordaron los aspectos legales de los datos, destacando su creciente importancia en la era digital y su comparación con el petróleo como recurso fundamental. Se subrayó la naturaleza única de los datos en cuanto a su reutilización, privacidad y capacidad para generar más datos. Además, se destacó el enorme volumen de datos generados y la necesidad de establecer regulaciones claras para garantizar su uso ético y responsable. Se planteó la importancia de mejorar el acceso y el intercambio de datos, así como la necesidad de un enfoque holístico para abordar los desafíos legales y tecnológicos asociados con los datos. En última instancia, se argumentó que el desbloqueo del valor de los datos es importante para impulsar la innovación y el desarrollo tecnológico en un entorno colaborativo, especialmente en el contexto de la IA.

Las teorías legales de los datos representan un campo complejo y en constante evolución en el que se abordan diversos aspectos relacionados con la propiedad, la privacidad, la competencia y los derechos del consumidor. Desde la perspectiva de la protección de datos personales hasta la regulación de la competencia en el mercado de datos, estas teorías ofrecen enfoques distintos para abordar los desafíos planteados por la era digital y la economía de los datos.

Es evidente que existe una discrepancia entre estas teorías, ya que cada una enfatiza diferentes aspectos y derechos en relación con los datos. Mientras que algunas se centran en proteger la privacidad individual y los derechos de los consumidores, otras buscan establecer derechos de propiedad sobre los datos o garantizar una competencia justa en el mercado. Sin embargo, todas estas teorías comparten el objetivo común de encontrar un equilibrio entre la protección de los intereses individuales y la promoción del bienestar general en el contexto de la economía de datos.

A medida que se avanza hacia un futuro cada vez más digitalizado, será decisivo seguir debatiendo y refinando estas teorías para adaptarlas a los nuevos desafíos y realidades que surgen en torno a los datos. Además, será importante considerar cómo estas teorías pueden converger y sinergizar para abordar de

manera efectiva los complejos problemas legales y éticos asociados con la gestión y el uso de los datos en la sociedad actual.

Ante la necesidad de una nueva teoría de datos, se evidencia la complejidad inherente a los derechos sobre los mismos. Las teorías existentes no logran abarcar todas las circunstancias pertinentes, lo que motiva la exploración de un nuevo enfoque. Surge así la noción de un derecho sobre los datos, que no se limita a los derechos de personalidad, secretos comerciales o propiedad intelectual, sino que representa una forma de propiedad incompleta con implicaciones tanto privadas como públicas.

Este nuevo paradigma reconoce la dualidad de los datos como fragmentos de comportamientos humanos y recursos sociales valiosos, requiriendo una combinación de derechos de personalidad y derechos de propiedad. Además, se destaca la necesidad de regular los derechos sobre los datos de manera integral, considerando aspectos como la producción, acceso, uso y disposición de los mismos.

Sin embargo, la implementación de esta teoría enfrenta desafíos tecnológicos y éticos significativos, como la seguridad de los datos y la protección de la privacidad. Se propone un enfoque colaborativo entre diversos actores para equilibrar los intereses públicos y personales, así como los intereses patrimoniales y de la personalidad.

Mirando hacia el futuro, se vislumbra un cambio paradigmático donde el derecho a los datos podría convertirse en un cuarto derecho humano fundamental, junto con el derecho a la vida, propiedad y libertad. Este nuevo enfoque requiere una comprensión profunda de la naturaleza cambiante de los datos y un marco legal y ético adaptable para garantizar su protección y utilización justa en la sociedad moderna.

El marco jurídico vigente para el acceso e intercambio de datos, especialmente en la UE, ha evolucionado notablemente desde la protección de datos personales hasta una visión más amplia que abarca diversos tipos de datos y teorías legales. A través de regulaciones como el RGPD, el Reglamento de Gobernanza de Datos y el Reglamento de Datos, la UE ha establecido un marco

legal integral que refleja la complejidad de los datos en la era digital, considerando su valor económico y social, así como los derechos y protecciones necesarios para individuos y empresas. Mientras tanto, otras naciones como México también están desarrollando su propia regulación, influenciados por las normativas europeas pero adaptándolas a su contexto nacional. Se espera que en el futuro se consideren aún más aspectos, como la IA en la regulación de los datos, que debe ser holística y estar en consonancia con el desarrollo tecnológico y las necesidades de la sociedad.

Asimismo, se abordó la complejidad de los datos y su clasificación legal en el contexto digital actual. Se destaca la necesidad de comprender la diversidad de los datos y las normativas que los regulan para abordar los desafíos éticos, sociales y legales. Se discuten diferentes marcos de clasificación y se propone una Teoría de Clasificación de Datos en Tres Niveles: Obligatorio, Buenas Prácticas y Consideraciones Futuras. Este enfoque estructurado ofrece claridad y orientación en la gestión de datos, permitiendo a las organizaciones priorizar y manejar los datos de manera más efectiva en un entorno cada vez más impulsado por la información.

En este sentido, se realizó un estudio exhaustivo de los distintos tipos de datos en el ámbito obligatorio revelando la complejidad y diversidad de las características que deben considerarse en la gestión y tratamiento de la información. Desde datos personales hasta su nivel de confidencialidad, cifrado y disponibilidad, cada aspecto demanda una atención meticulosa para garantizar la protección de la privacidad y la seguridad de los datos, así como su adecuado uso y manejo en el contexto de productos y servicios relacionados.

La teoría de clasificación de datos en tres niveles proporciona un marco integral para entender la diversidad y complejidad de los datos en el ámbito de la ciencia de datos y la IA. Desde la categorización de datos personales y no personales hasta la consideración de datos mentales y sintéticos en el futuro, este enfoque aborda aspectos vitales como la seguridad, la privacidad y la calidad de los datos. A medida que se avanza hacia consideraciones futuras, la atención se

centra en la naturaleza efímera y fragmentada de los datos, lo que subraya la necesidad de adaptación continua en este campo en evolución constante.

Asimismo, se enlistaron todos los actores y roles en los datos. Específicamente se ha trazado un panorama completo de los diversos actores y roles involucrados en el ecosistema de datos. Desde el Interesado, quien es el sujeto central cuyos datos están siendo procesados, hasta los distintos profesionales como el Delegado de Protección de Datos, el *Data Scientist*, y el *Data Analyst*, quienes trabajan en la manipulación, análisis y aplicación de los datos. Además, se abordaron roles de liderazgo y toma de decisiones, como los Oficiales/Directores vinculados a los datos y los comités de gobernanza de datos, que desempeñan un papel clave en la formulación de políticas y estrategias para garantizar el cumplimiento normativo y la gestión efectiva de los datos. Este mapeo exhaustivo de roles subraya la complejidad y la interdependencia de los actores dentro del entorno de los datos, destacando la necesidad de una colaboración cohesionada y una asignación clara de responsabilidades para maximizar el valor de los datos y mitigar los riesgos asociados con su uso.

Finalmente, se exploraron diversos enfoques para mejorar el acceso y el intercambio de datos, abarcando desde descargas tradicionales hasta tecnologías más avanzadas como las APIs, los entornos de prueba (*sandbox*) y las tecnologías para mejorar la privacidad (PETs). Estas herramientas ofrecen soluciones variadas para optimizar la manera en que se accede y se comparten datos, brindando opciones adaptativas a las necesidades específicas de cada contexto y priorizando la seguridad y la privacidad.

Y en el mismo sentido, se exploró una amplia gama de enfoques y modelos para el intercambio de datos, que van desde transferencias simples hasta estructuras más complejas como mercados de datos, corredores de datos, intercambios de datos y más. Estos conceptos representan diversas formas de organizar y facilitar la colaboración y el intercambio de datos entre diferentes entidades, ya sean individuos, empresas u organizaciones. Al comprender estas diferentes modalidades, se puede diseñar estrategias efectivas para gestionar

datos de manera colaborativa, asegurando la protección de la privacidad, la seguridad y promoviendo el beneficio mutuo entre las partes involucradas.

Así como examinar los diferentes tipos de interacciones con los datos, destacando dos aspectos fundamentales: el uso de datos para alimentar sistemas y su transmisión y transferencia. Estas interacciones reflejan la diversidad de funciones y procesos en los cuales los datos desempeñan un papel crucial, ya sea como fuente de información para alimentar sistemas inteligentes o como medio para compartir información entre diferentes entidades. Comprender estos tipos de interacciones es esencial para optimizar la gestión y el flujo de datos, garantizando su uso efectivo y seguro en diversas aplicaciones y contextos.

En resumen, el análisis exhaustivo realizado abordó la complejidad legal y ética que rodea el mundo de los datos en la era digital. Desde la comparación con el petróleo como recurso esencial hasta la propuesta de un nuevo paradigma de derechos sobre los datos, se destaca la necesidad de regulaciones claras y adaptativas. Además, se propone una Teoría de Clasificación de Datos en Tres Niveles para abordar la diversidad y complejidad de los datos, junto con un mapeo detallado de los diversos actores y roles en el ecosistema de datos. Se exploran también múltiples enfoques para mejorar el acceso y el intercambio de datos, priorizando la seguridad, la privacidad y la colaboración efectiva. Estas reflexiones son fundamentales para adaptarse a los desafíos cambiantes y aprovechar el potencial de los datos en la sociedad moderna.

Bibliografía

Abrams, M. (2014). The Origins of Personal Data and its Implications for Governance, *The Information Accountability Foundation,* https://ssrn.com/abstract=2510927

Access Now. (2018). *Mapping Regulatory Proposals for Artificial Intelligence in Europe*. https://www.accessnow.org/cms/assets/uploads/2018/11/mapping_regulatory_proposals_for_AI_in_EU.pdf

Ada Lovelace Institute. (2020). *Examining the Black Box: Tools for assessing algorithmic systems Identifying common language for algorithm audits and impact assessments*. https://bit.ly/3vmItGB

Adams, E. (2024). *See Yourself in Cyber: Security Careers Beyond Hacking*. John Wiley & Sons.

ACLU. (2022). *ACLU v. Department of Homeland Security (commercial location data FOIA)*, https://www.aclu.org/cases/aclu-v-department-homeland-security-commercial-location-data-foia

AEPD. (2019). *A Guide to Privacy by Design*. https://www.aepd.es/es/documento/guia-privacidad-desde-diseno_en.pdf

AEPD. (2020a). *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción*. https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

AEPD. (2020b). *Guidelines for Data Protection by Default*. https://www.aepd.es/sites/default/files/2020-10/guia-proteccion-datos-por-defecto-en.pdf

AEPD. (2020c). *GDPR compliance of processings that embed Artificial Intelligence An introduction*. https://www.aepd.es/guides/gdpr-compliance-processings-that-embed-ia.pdf

Antonopoulos, A. & Wood, G. (2019). *Mastering Ethereum Building Smart Contracts and Dapps*. O'Reilly.

Arora, B. (2018) Chapter 2. Big Data Analytics: The Underlying Technologies Used by Organizations for Value Generation, En H. Chahal, J. Jyoti & J. Wirtz (eds), *Understanding the Role of Business Analytics: Some Applications* (9-30). Springer.

Ballardini, R. et al. (2019). *Regulating Industrial Internet Through IPR, Data Protection and Competition Law*. Kluwer Law International.

Bambara, J. & Allen, P. (2018). *Blockchain A Practical Guide to Developing Business, Law, and Technology Solutions*. McGraw-Hill.

Banco Mundial. (2021). *World Development Report 2021: DATA FOR BETTER LIVES*. https://openknowledge.worldbank.org/bitstream/handle/10986/35218/9781464816000.pdf

Banterle, F. (2018). The Interface Between Data Protection and IP Law: The Case of Trade Secrets and the Database sui generis Right in Marketing Operations, and the Ownership of Raw Data in Big Data Analysis. En M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 411-443). Springer. https://doi.org/10.1007/978-3-662-57646-5_16

Barocas, S. (2014). Data mining and the discourse on discrimination. *Center for Information Technology Policy Princeton University*.

Bashir, I. (2020). *Mastering Blockchain A deep dive into distributed ledgers, consensus protocols, smart contracts, DApps, cryptocurrencies, Ethereum, and more*, 3ra ed. Packt Publishing.

Bendiek, S. (2017). “Preface: the “Cloud Way” to Digital Transformation and New Business Models”. En C. Linnhoff-Popien, R. Schneider, & M. Zaddach (eds.), *Digital Marketplaces Unleashed* (pp. 691-698). Springer.

Bollweg, L. (2022). *Data Governance for Managers: The Driver of Value Stream Optimization and a Pacemaker for Digital Transformation.* Springer Nature.

Burdon, M. (2020). *Digital Data Collection and Information Privacy Law.* Cambridge University Press.

Burns, B. (2018). Designing Distributed Systems Patterns and Paradigms for Scalable, Reliable Services. O'Reilly.

Butterfield, A. et al. (eds). (2016). *A Dictionary of Computer Science*, 7ma ed. Oxford University Press.

California Consumer Privacy Act. (2018). TITLE 1.81.5. California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100], https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

Cameron, F. (2021). *The Future Of Digital Data, Heritage And Curation in a More-than-Human World*. Routledge.

Cao, L. (2018). *Data Science Thinking The Next Scientific, Technological and Economic Revolution*. Springer International Publishing.

Carruthers, C. & Jackson, P. (2021). *The Chief Data Officer's Playbook,* 2da ed. Facet Publishing.

Chapple, M. et al. (2018). *CISSP Certified Information Systems Security Professional Official Study Guide*, 8va ed. Sybex.

Chapple, M. & Shelley, J. (2021). *IAPP CIPP/US Certified Information Privacy Professional Study Guide United States Exam*. Sybex.

Charles, V. & Emrouznejad, A. (2019). 'Chapter 1 Big Data for the Greater Good: An Introduction', En A. Emrouznejad & V. Charles (eds), *Big Data for the Greater Good* (pp. 1-18). Springer.

Chirita, A. (2018). The Rise of Big Data and the Loss of Privacy, In M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 153-189). Springer. https://doi.org/10.1007/978-3-662-57646-5_7

Chojecki, P. (2020). Data Science Job: How to become a Data Scientist. Independently published.

Christen, P. et al. (2020). *Linking Sensitive Data Methods and Techniques for Practical Privacy-Preserving Information Sharing*. Springer.

Chrobak, L. (2018). Proprietary Rights in Digital Data? Normative Perspectives and Principles of Civil Law. In M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 253-272). Springer. https://doi.org/10.1007/978-3-662-57646-5_10

Church, P., & Cumbley, R. (2022). 10 DATA AND DATA PROTECTION. En C. Kerrigan (ed.), *ARTIFICIAL INTELLIGENCE Law and Regulation* (pp. 163-196). Edward Elgar Publishing.

Cielen, D. et al. (2016). *Introducing Data Science Big Data, Machine Learning, And More, Using Python Tools*. Manning Publications.

Cohen, J. (2019). *Between Truth and Power The Legal Constructions of Informational Capitalism*. Oxford University Press.

Collmann, J., & Matei, S. A. (2016). Ethical reasoning in big data: An exploratory analysis. Springer. https://doi.org/10.1007/978-3-319-28422-4_1

Comisión Europea. (s.f.). A European strategy for data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data

Comisión Europea. (2017). Propuesta de Reglamento del Parlamento Europeo y del Consejo sobre el respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas y por el que se deroga la Directiva 2002/58/CE (Reglamento sobre la privacidad y las comunicaciones electrónicas). COM(2017) 10 final 2017/0003(COD). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017PC0010

Comisión Europea. (2019). *European Commission Cloud Strategy Cloud as an enabler for the European Commission Digital Strategy 16 May 2019 V.1.0.1*. https://commission.europa.eu/document/download/3bb440ec-f777-484c-8802-baf08ebb87c0_en?filename=ec_cloud_strategy.pdf

Comisión Europea. (2020a). *Una Estrategia Europea de Datos*. https://eur-lex.europa.eu/legal-content/ES/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066

Comisión Europea. (2020b). *WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust*. Brussels, 19.2.2020 COM(2020) 65. https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Committee on Government Operations (EE. UU.). (1975). Privacy and Protection of Personal Information in Europe: Privacy Developments in Europe and Their Implications for United States Policy: a Staff Report of the Committee on Government Operations, United States Senate. Government Printing Office.

Concilio, G., & Pucci, P. (2021). Part I The Data Shake: Open Questions and Challenges for Policy Making. In G. Concilio, P. Pucci, L. Raes, & G. Mareels (Eds.), *The Data Shake Opportunities and Obstacles for Urban Policy Making* (pp. 03-18). Springer. https://doi.org/10.1007/978-3-030-63693-7

Consejo de Europa (1981). Convenio para la Protección de las Personas con Respecto al Tratamiento Automatizado de Datos de Caracter Personal. https://rm.coe.int/16806c1abd

Consejo de la Unión Europea. (2020). Council Resolution on Encryption - Security through encryption and security despite encryption. https://data.consilium.europa.eu/doc/document/ST-13084-2020-REV-1/en/pdf

Constitución Política de la República de Chile. (1980). Constitución Política De La República (Texto Actualizado Al 22 De Enero De 2024). https://www.camara.cl/camara/doc/leyes_normas/constitucion.pdf

Corea, F. (2019). *An Introduction to Data Everything You Need to Know About AI, Big Data and Data Science*. Springer.

Curry, E. et al. (2022) 'Data Spaces: Design, Deployment, and Future Directions', En E. Curry et al. (ed.), *Data Spaces Design, Deployment and Future Directions* (pp. 01-17). Springer.

Dale, N., & Lewis, J. (2016). *Computer Science Illuminated* (6th ed.). Jones & Bartlett Learning.

Datatilsynet (2015). *A guide to the Anonymisation Of Personal Data*. https://www.datatilsynet.no/link/2e642d84d9214490866a297a71a44c78.aspx/download

Datoo, A. (2019). *Legal Data for Banking: Business Optimisation and Regulatory Compliance*. Wiley.

Davara, I. et al. (2019) Titular de los datos personales, En I. Davara F. de Marcos (ed.), *Diccionario de Protección de Datos Personales Conceptos Fundamentales* (pp. 850-853). INAI.https://home.inai.org.mx/wp-content/documentos/Publicaciones/Documentos/DICCIONARIO_PDP_digital.pdf

Davidof, S. (2020). *Data Breaches Crisis and Opportunity*. Addison-Wesley.

Davis, K. (2012). Ethics of Big Data: Balancing Risk and Innovation. Sebastopol, CA: O'Reilly Media.

De Guise, P. (2020). *Data Protection Ensuring Data Availability,* 2da ed. Auerbach Publications. https://doi.org/10.1201/9780367463496

Determann, L. (2022). Determann's Field Guide to Data Privacy Law International Corporate Compliance, 5ta ed., Edward Elgar Publishing.

Diario Oficial de la Unión Europea. (2024a, 17 de diciembre). Corrección de errores del Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo, de 13 de diciembre de 2023, sobre normas armonizadas para un acceso justo a los datos y su utilización, y por el que se modifican el Reglamento (UE) 2017/2394 y la Directiva (UE) 2020/1828 (Reglamento de Datos). Serie L, 2024/90814. Recuperado de http://data.europa.eu/eli/reg/2023/2854/corrigendum/2024-12-17/oj

Diario Oficial de la Unión Europea. (2024b, 17 de diciembre). Corrección de errores del Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo, de 30 de mayo de 2022, relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos). Serie L, 2024/90815. Recuperado de http://data.europa.eu/eli/reg/2022/868/corrigendum/2024-12-17/oj

Dignum, V. (2019). *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*. Springer.

Directiva 2002/58/CE del Parlamento Europeo y del Consejo, de 12 de julio de 2002, relativa al tratamiento de los datos personales y a la protección de la

intimidad en el sector de las comunicaciones electrónicas (Directiva sobre la privacidad y las comunicaciones electrónicas). Diario Oficial de la Unión Europea, L 201, de 31 de julio de 2002. http://data.europa.eu/eli/dir/2002/58/oj

Directiva (UE) 2019/1024 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. Diario Oficial de la Unión Europea, L 172/56, de 20 de junio de 2019. http://data.europa.eu/eli/dir/2002/58/oj

DOF (Diario Oficial de la Federación). (2018). DOF: 28/09/2018 DECRETO Promulgatorio del Convenio para la Protección de las Personas con respecto al Tratamiento Automatizado de Datos de Carácter Personal, hecho en Estrasburgo, Francia, el veintiocho de enero de mil novecientos ochenta y uno. https://www.dof.gob.mx/nota_detalle.php?codigo=5539473&fecha=28/09/2018#gsc.tab=0

Doshi, H. (2022). *Certified Information Security Manager Exam Prep Guide*, 2da ed. Packt Publishing Ltd.

EDPB-EDPS. (2021). EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) 18 June 2021. https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf

EDPB-EDPS. (2022a). EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on theEuropean Health Data Space Adopted on 12 July 2022. https://www.edpb.europa.eu/system/files/2022-07/edpb_edps_jointopinion_202203_europeanhealthdataspace_en.pdf

EDPB-EDPS. (2022b). EDPB-EDPS Joint Opinion 2/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) Adopted on 4 May 2022. https://www.edpb.europa.eu/system/files/2022-05/edpb-edps_joint_opinion_22022_on_data_act_proposal_en.pdf

ENISA. (2021). *Data Pseudonymisation: Advanced Techniques and Use Cases*. https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases

European Union Agency for Fundamental Rights. (2020). Getting the future right artificial intelligence and fundamental rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

Executive Office of the President. (2024). M-24-10 memorandum for the heads of executive departments and agencies, March 28, 2024. https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

Fenwick, M et al. (2018). Business and Regulatory Responses to Artificial Intelligence: Dynamic Regulation, Innovation Ecosystems and the Strategic Management of Disruptive Technology. En M. Corrales, M. Fenwick y N. Forgó, (eds.), *Robotics, AI and the Future of Law*. (pp. 81-103) Springer.

Ferrer, E. (2019) 'Intimidad', En I. Davara F. de Marcos (ed.), *Diccionario de Protección de Datos Personales Conceptos Fundamentales* (pp. 459-464). INAI.https://home.inai.org.mx/wp-content/documentos/Publicaciones/Documentos/DICCIONARIO_PDP_digital.pdf

Figueroa, G. (1998). Los derechos de la personalidad en general: concepción tradicional. *Revista de Derecho de la Universidad Católica de Valparaíso* (19) 21-33. https://www.projurepucv.cl/index.php/rderecho/article/download/397/370/0

Fitzgibbons, L. (2019). states of digital data, *Techtarget*. https://searchdatamanagement.techtarget.com/reference/states-of-digital-data

Fleckenstein, M & Fellows, L., (2018). Modern Data Strategy. Springer.

Fox, D. (2011). '17 The Right to Silence Protects Mental Control', En M. Freeman (ed.), *Law and Neuroscience* (pp. 335-365). Oxford University Press.

FRA. (2019). *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*.

https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf

Fryman, L. et al. (2016). *The Data and Analytics Playbook: Proven Methods for Governed Data and Analytic Quality*. Morgan Kaufmann.

G20. (2019). *G20 Osaka Leaders' Declaration*. https://www.consilium.europa.eu/media/40124/final_g20_osaka_leaders_declaration.pdf

Garfinkel, S. (2015). *NISTIR 8053 De-Identification of Personal Information*, https://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf

GDC NCI. (s.f.). Accessing Genomic Data. https://gdc.cancer.gov/access-data

Ghavami, P. (2020). *Big Data Management: Data Governance Principles for Big Data Analytics*. Walter de Gruyter GmbH & Co KG.

Gilchrist, A. (2016). *Industry 4.0 The Industrial Internet of Things*. Apress.

Graef, I. (2018). Blurring Boundaries of Consumer Welfare How to Create Synergies Between Competition, Consumer and Data Protection Law in Digital Markets. En M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 121-151). Springer. https://doi.org/10.1007/978-3-662-57646-5_6

Greenleaf, G. (2019). G20 Makes Declaration of 'Data Free Flow With Trust': Support and Dissent. *Privacy Laws & Business International Report*, (160), 3. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3514407

Gregory, P. (2018). *CISM® Certified Information Security Manager Exam Guide*. McGraw-Hill.

Gobierno de España. (2021). 'Carta de Derechos Digitales'. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf

Gordon, A. (2016). The Official (ISC)2 Guide to the CCSP CBK. Wiley.

Hafen, e. (2019). Personal Data Cooperatives - A New Data Governance Framework for Data Donations and Precision Health. En J. Krutzinna & L. Floridi (eds.). *The Ethics of Medical Data Donation* (141-150). Springer.

Hamon, R. et al. (2020). *Robustness and Explainability of Artificial Intelligence, European Commission JRC Technical Report.* Publications Office of the European Union.

Han, J. et al. (2012). *Data Mining Concepts and Techniques*, 3ra ed. Elsevier.

Harjika, R. (2022). *Realize Enterprise Architecture with AWS and SAFe: A comprehensive, hands-on guide to AWS with Agile and TOGAF*. Packt Publishing Ltd.

Haverbeke, M. (2018). *Eloquent JavaScript*, 3ra ed. No Starch Press.

Heerde, V., et al. (2009). "A Framework to Balance Privacy and Data Usability Using Data Degradation", *2009 International Conference on Computational Science and Engineering*. https://doi.org/10.1109/CSE.2009.174

Hirsch, D. et al. (2023). *Business Data Ethics: Emerging Models for Governing AI and Advanced Analytics*. Springer Nature.

Hollifield, A. & Cofey, A. (2023). *Media Analytics Understanding Media, Audiences, and Consumers in the 21st Century*. Routledge.

Holt, A. (2021). *Data Governance Governing data for sustainable business*. BCS Learning and Development Ltd.

Hopper, M. (2023). *Practitioner's Guide to Operationalizing Data Governance*. Wiley.

Hurwitz, J. et al. (2016) Big Data for Dummies. Wiley.

Hurwitz, J. et al. (2020). *Augmented Intelligence the business power of human–machine collaboration*. CRC Press.

IAPP. (s.f.). *Data Aggregation*. Glossary of Privacy Terms. https://iapp.org/resources/glossary

IBM. (2024). *Cost of a Data Breach Report 2024*. https://www.ibm.com/downloads/documents/us-en/107a02e94948f4ec

ICO. (s.f.). *The Guide to the Sandbox*. https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/the-guide-to-the-sandbox/

ICO. (2012). *Anonymisation: managing data protection risk code of practice*. https://ico.org.uk/media/1061/anonymisation-code.pdf

ICO. (2013). *Guidance on the use of cloud computing*. (https://ico.org.uk/media/about-the-ico/documents/1042330/cloud-computing-guidance-for-organisations.pdf

ICO. (2021). *Guide to the General Data Protection Regulation (GDPR)*. https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr-1-1.pdf

IDC & Seagate. (2018). *Data Age 2025 The Digitization of the World From Edge to Core*. https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Isson, J. (2018). *Unstructured Data Analytics How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention.* Wiley.

IT Governance Privacy Team. (2020). An implementation and compliance guide, 4ta ed. IT Governance Publishing.

Jägare, U. (2019). *Data Science Strategy.* John Wiley & Sons, Inc.

Jarmul, K. (2023) *Practical Data Privacy Solving Privacy and Security Problems in Your Data Science Workflow*, O'Reilly Media.

Jasmontaitė-zaniewicz, L., et. al. (2021). *The GDPR Made Simple(R) For SMES*. VUBPRESS.

Kelleher, J. D., & Tierney, B. (2018). *Data science*. The MIT Press. https://doi.org/10.7551/mitpress/11140.001.0001

Kelsey, F. (2016). *Future Of Privacy Forum. A Visual Guide to Practical Data De-Identification*. https://fpf.org/blog/a-visual-guide-to-practical-data-de-identification/

Kim, D. & Solomon, M. (2021). *Fundamentals of Information Systems Security*, 4ta ed., Jones & Bartlett Learning.

King, T. & Schwarzenbach, J. (2020). *Managing Data Quality A practical guide.* BCS Learning & Development Ltd.

Klidas, A. & Hanegan, K. (2022). *Data Literacy in Practice: A complete guide to data literacy and making smarter decisions with data through intelligent actions*. Packt Publishing Ltd.

Kitchin. R. (2014). T*he Data Revolution Bid Data, Open Data, Data Infraestructures & Their Consequences*. Sage.

Koene, A. et al. (2019). *A governance framework for algorithmic accountability and transparency*. European Parliamentary Research Service.

Krishnakumar, A. (2020). *Quantum Computing and Blockchain in Business Exploring the applications, challenges, and collision of quantum computing and blockchain.* Packt Publishing.

Lambert, P. (2022) *Gringras: The Laws of the Internet.* Bloomsbury Publishing.

Llamas Covarrubias, J. Z. (2020). La propiedad como objeto, medio y fin en el uso de las tic: propiedades inteligentes, geoetiquetadas y virtuales. *Revista de la propiedad inmaterial diciembre 2020-junio 2021*, (30), 247-293. https://doi.org/10.18601/16571959.n30.10

Llamas Covarrubias, et al. (2022). Enfoques regulatorios para la Inteligencia Artificial (IA). Revista Chilena De Derecho, 49(3), 31–62. https://doi.org/10.7764/R.493.2

Lesk, A. (2014). Introduction to Bioinformatics, 4ta ed. Oxford University Press.

Leslie, D. & Briggs, M. (2021). *Explaining Decisions Made With AI: A Workbook Use Case 2: Machine Learning for Children's Social Care*. The Alan Turing Institute.

Liu, A. & Li R. (2021). *Algorithms for Data and Computation Privacy*. Springer.

Lundqvist, B. (2018). Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet-of-Things World: The Issue of Accessing Data. En M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 191-214). Springer.

Lundqvist, B. (2023). *Regulating Access and Transfer of Data*. Cambridge University Press.

Lyn, J. (2020). *Legal and Privacy Issues in Information Security*, 3ra ed. Jones & Bartlett Learning.

Macmillan, M. (2018). Chapter 4: Data Protection Concepts. En E. Ustaran, & H. Lovells (eds.), *European Data Protection Law and Practice, 2da ed*. IAPP.

Mahanti, R. (2019). *Data Quality: Dimensions, Measurement, Strategy, Management, and Governance*. Quality Press.

Maqueo, M. et al. (2021). Evaluación del grado de Preparación para la adopción de Inteligencia Artificial en los sistemas judiciales de la Región caso México. https://bit.ly/3o0gzO2.

Marr, B. (2017). *Data Strategy How to profit from a world of big data, analytics and the internet of things.* Kogan Page Limited.

Mattsson, U. (2022). *Controlling Privacy and the Use of Data Assets Volume 1 Who Owns the New Oil?*. CRC Press.

Martens, D. (2022). Data Science Ethics: Concepts, Techniques, and Cautionary Tales. Oxford University Press.

Martin, D. et al. (2024). Enhancing collabortative enuroimaging research: introducing COINSTAC Vaults for federated analysis and reproducibility. En M. Van Swieten & C. Haselgrove (Eds.), *Nabigating the landscape of FAIR data sharing and reuse: Repositories, standards, and resources* (pp. 39-51). Frontiers.

Martin, K. (2022). Ethics of data and analytics: Concepts and cases. CRC Press.

Mcintosh, V. (2019). *Understanding aggregate, de-identified and anonymous data*, Comparitech. https://www.comparitech.com/blog/information-security/aggregate-vs-anonymous-data/

Miao, F., et al. (2022). “Data Architecture for Big Data Service Operations Management (The New Vision of Data Architecture for the Future Human Society)”. En A. Emrouznejad & V. Charles (eds.), *Big Data and Blockchain for Service Operations Management* (pp. 95-137). Springer.

Microsoft. (2014). Real-Time Data: Frequently Asked Questions. https://docs.microsoft.com/en-us/previous-versions/office/developer/office-xp/aa140060(v=office.10)?redirectedfrom=MSDN

Montes, L. (2020) “Protocolo Para El Tratamiento De Información De Carácter Personal En El Marco Del Programa De Donación De Cuerposde La Facultad De Medicina De La UNAM”, [Tesis de maestría]. INFOTEC. https://infotec.repositorioinstitucional.mx/jspui/bitstream/1027/388/1/INFOTEC_MDTIC_LMB_17072010.pdf

Mortazavi, M. & Khaled, S. (2015) ‘Chapter 3 Privacy and Big Data’, En S. Zeadally & M. Badra (eds.), *Privacy in a Digital, Networked World Technologies, Implications and Solutions* (pp. 37-55). Springer.

Moschovitis, C. (2021). *Privacy, Regulations, And Cybersecurity The Essential Business Guide.* Wiley.

Mueller, J. & Massaron, L. (2019). *Data Science Programming All-in-One For Dummies.* Wiley.

Muller, C. (2020). Chapter I. The Impact of AI on Human Rights, Democracy and the Rule of Law. In *Towards Regulation of AI Systems. Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe’s standards on human rights, democracy and the rule of law* (pp. 26-41). Comité ad hoc de Inteligencia Artificial (Unión Europea).

Munir K. & Mohammed, L. (2019) Chapter 36. Cloud Computing Data Storage Security Based on Different Encryption Schemes, En Management Association, Information Resources (ed), *Cloud Security: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications* (pp. 712-743). IGI Global.

Narayanan, A. & Shmatikov, V. (2008). *Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset)*, University of Texas at Austin. https://arxiv.org/pdf/cs/0610105.pdf

Nikolinakos, N. (2023). *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies - The AI Act*. Springer.

NIST. (s.f.) *Computer Security Resource Center, data confidentiality*. https://csrc.nist.gov/glossary/term/confidentiality

NIST. (s.f.) *Computer Security Resource Center, Non-Volatile Data*. https://csrc.nist.gov/glossary/term/non_volatile_data

NIST. (s.f.) *Computer Security Resource Center, Volatile Data*. https://csrc.nist.gov/glossary/term/volatile_data

NIST. (s.f.) *Computer Security Resource Center, perishable data*. https://csrc.nist.gov/glossary/term/perishable_data

NIST. (s.f.) *Computer Security Resource Center, RED data*. https://csrc.nist.gov/glossary/term/red_data

NIST. (s.f.) *Computer Security Resource Center, BLACK data*. https://csrc.nist.gov/glossary/term/black_data

NIST. (s.f.) *Computer Security Resource Center, obscured data*. https://csrc.nist.gov/glossary/term/obscured_data

NIST. (s.f.) *Computer Security Resource Center, classified information*. https://csrc.nist.gov/glossary/term/classified_information

NIST. (s.f.) *Computer Security Resource Center, proscribed information*. https://csrc.nist.gov/glossary/term/proscribed_information

NIST. (s.f.) *Computer Security Resource Center, restricted data*. https://csrc.nist.gov/glossary/term/restricted_data

NIST. (s.f.) *Computer Security Resource Center, de-identified information*. https://csrc.nist.gov/glossary/term/de_identified_information

OCDE. (2015). *OECD Digital Government Studies Open Government Data Review of Poland Unlocking the Value of Government Data: Unlocking the Value of Government Data*. OECD Publishing. https://doi.org/10.1787/9789264241787-en

OCDE. (2019a). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*. OECD Publishing. https://doi.org/10.1787/276aaca8-en

OCDE. (2019b). *The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies*. OECD Publishing. https://doi.org/10.1787/059814a7-en

OCDE. (2021). *State of Implementation of The OECD AI Principles Insights from National AI Policies*. OECD Publishing. https://doi.org/10.1787/1cd40c44-en

OEA. (2019). *Clasificación de Datos White paper series Edición 6*. https://www.oas.org/es/sms/cicte/docs/ESP-Clasificacion-de-Datos.pdf

O'Keefe, K., & O'Brien, D. (2018). Ethical data and information management: Concepts, tools and methods. London: Kogan Page.

ONU. (2020). *Data Strategy of the Secretary-General for Action by Everyone, Everywhere with Insight, Impact and Integrity 2020-22*. https://www.un.org/en/content/datastrategy/images/pdf/UN_SG_Data-Strategy.pdf

Open Data Handbook. (s.f.) 'Glossary'. https://opendatahandbook.org/glossary/en/

Overly, M. (2021). *A Guide to IT Contracting Checklists, Tools, and Techniques*, 2da ed. CRC Press.

Parlamento Europeo. (2020). *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. EPRS | European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf

Parra-Moyano, J. (2021). 3 Shared Data: Backbone of a New Knowledge Economy. En A. Pentland, A. Lipton & T. Hardjono (Eds.), *Building The New Economy Data As Capital* (pp. 35-42). The MIT Press.

Pasquale, F. (2015). *The Black Box Society The Secret Algorithms That Control Money and Information*. Harvard University Press.

PDPC. (2021). *Advisory Guidelines On Key Concepts In The Personal Data Protection Act*. https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Key-Concepts/Advisory-Guidelines-on-Key-Concepts-in-the-PDPA-1-Oct-2021.pdf?la=en

Ploeger, H. & Van Loenen, B. (2018) 'Chapter 14 2050: The Story of Urbidata', En BV. Loenen, G. Vancauwenberghe, J. Crompvoets (eds), *Open Data Exposed* (pp. 269-288). Springer.

Plotkin, D. (2021). *Data Stewardship An Actionable Guide to Effective Data Management and Data Governance*, 2da ed. Academic Press Elsevier.

Plotkin, D. (2023). Human Resources Management and Data Governance Roles: Executive Sponsor, Data Governors, and Data Stewards, En I. Caballero & M.

Piattini (Eds.), *Data Governance From the fundamentals to real cases* (pp. 57-74). Springer.

Pierson, L. (2021). *Data Science for dummies*, 3ra ed. John Wiley & Sons, Inc.

Preukschat, A. & Reed, D. (2021). *Self-Sovereign Identity decentralized digital identity and verifiable credentials*. Manning Publications Co.

Ratti, M. (2018). Personal-Data and Consumer Protection: What Do They Have in Common?, En M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 377-393). Springer.

Raj, P. & Raman, A. (2017). *The Internet of Things Enabling Technologies, Platforms, and Use Cases.* CRC Press.

Rayes, A., & Salam, S. (2019). *Internet of Things From Hype to Reality The Road to Digitization*, 2da edición, Springer.

REDIPD. (2019). Orientaciones Específicas para el Cumplimiento de los Principios y Derechos que Rigen la Protección de los Datos Personales en los Proyectos de Inteligencia Artificial. https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-espec%C3%ADficas-proteccion-datos-ia.pdf

REDIPD. (2019). Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial. RIPDD. https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf

Reed, C. (2021). 5 Information Ownership in the Cloud. In C. Millard (Ed.), *Cloud Computing Law* (2da ed., pp. 175-207). Oxford University Press. https://doi.org/10.1093/oso/9780198716662.003.0005

Reglamento (CE) 2016/679 de la Comisión, 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 (Reglamento general de protección de datos). Diario Oficial

de la Unión Europea, L119/1, de 27 de abril de 2016. http://data.europa.eu/eli/reg/2016/679/oj

Reglamento (CE) 2018/1807 de la Comisión, de 14 de noviembre de 2018, relativo a un marco para la libre circulación de datos no personales en la Unión Europea. Diario Oficial de la Unión Europea, L303/59, de 14 de noviembre de 2018. http://data.europa.eu/eli/reg/2018/1807/oj

Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos). Diario Oficial de la Unión Europea, L 152/1, de 30 de mayo de 2022. http://data.europa.eu/eli/reg/2022/868/oj

Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828 (Reglamento de Mercados Digitales). Diario Oficial de la Unión Europea, L 265/1, de 14 de septiembre de 2022. http://data.europa.eu/eli/reg/2022/1925/oj

Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales). Diario Oficial de la Unión Europea, L 277/1, de 19 de octubre de 2022. http://data.europa.eu/eli/reg/2022/2065/oj

Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo de 13 de diciembre de 2023 sobre normas armonizadas para un acceso justo a los datos y su utilización, y por el que se modifican el Reglamento (UE) 2017/2394 y la Directiva (UE) 2020/1828 (Reglamento de Datos). Diario Oficial de la Unión Europea, Series L, de 13 de diciembre de 2023. https://eur-lex.europa.eu/eli/reg/2023/2854/oj

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.o

300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea, Series L, de 13 de junio de 2024. http://data.europa.eu/eli/reg/2024/1689/oj

Reichental, J. (2023). *Data Governance For Dummies®*. John Wiley & Sons, Inc.

Renieris, E. (2023). *Beyond Data: Reclaiming Human Rights at the Dawn of the Metaverse*. MIT Press.

Richter, H. (2018). The Power Paradigm in Private Law Towards a Holistic Regulation of Personal Data. In M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 527-577). Springer. https://doi.org/10.1007/978-3-662-57646-5_19

Riemann, R. (s.f.). Synthetic Data. European Data Protection Supervisor.

Sattler, A. (2018). From Personality to Property? Revisiting the Fundamentals of the Protection of Personal Data. In M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Personal Data in Competition, Consumer Protection and Intellectual Property Law Towards a Holistic Approach?* (pp. 27-54). Springer. https://doi.org/10.1007/978-3-662-57646-5_3

Schmarzo, B. (2016). *Big Data MBA Driving Business Strategies with Data Science*. Wiley.

Schreider, T. (2020). *Cybersecurity Law, Standards and Regulations*, 2da ed., Rothstein Publishing.

Sensen, Z. et al. (2024). “Data Operation and Management Practices in Ping an Group”. En J. Feng, F. Jiang, M. Luo & L. Zhang (Eds.), *Edge Computing – EDGE 2023 7th International Conference, Held as Part of the Services Conference Federation, SCF 2023 Shenzhen, China, December 17-18, 2023, Proceedings* (pp. 82-91). Springer.

Shah, C. (2020). *A Hands-On Introduction to Data Science*. Cambridge University Press. DOI: 10.1017/9781108560412

Shaheen, M. (2021) *Hybrid Encryption Algorithms over Wireless Communication Channels*. CRC Press.

Sharma, A & Pandey, H. (2020). 'Big Data and Analytics in Industry 4.0', En A. Nayyar & A. Kumar (eds), *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development* (pp. 57-72). Springer.

Simbiose Ventures Team. (2020). *Getting Started with Data The first book you should read to successfully get along with data*. Lean Publishing.

Sinha, V. C., Gupta, A., & Saxena, J. K. (2021). *Business Statistics*. SBPD Publications.

Smith, B. (2020). *Data Science Advanced Guide to Learn the Realms of Data Science Effectively*. Independently published.

Solano, J. et al. (2022). *Governing data and artificial intelligence for all Models for sustainable and just data governance.* European Parliamentary Research Service

Sozonov, V. (2020). *The Fundamentals of Data Science Big Data, Deep Learning, and Machine Learning: What you need to know about data science and why it matters.* Vinco Publishing.

Steinrötter, B. (2020). 10 The (Envisaged) Legal Framework for Commercialisation of Digital Data within the EU Data Protection Law and Data Economic Law As a Conflicted Basis for Algorithm-Based Products and Services. In M. Ebers & S. Navas (Eds.), *Algorithms and Law* (pp. 269-297). Cambridge University Press.

Swan, M. et al. (2020). *Quantum computing: physics, blockchains, and deep learning smart networks*. World Scientific Publishing Europe.

Tang, A. (2023). *Privacy in Practice Establish and Operationalize a Holistic Data Privacy Program.* CRC Press.

TechTarget. (s.f.) 'fixed data (permanent data, reference data, archival data, or fixed-content data)'. https://searchdatamanagement.techtarget.com/definition/fixed-data

Techterms. (s.f.) 'Output'. https://techterms.com/definition/output

The Law Society. (2023). *Blockchain: Legal & Regulatory Guidance*, 3ra ed. https://prdsitecore93.azureedge.net/-/media/files/topics/research/blockchain-rep

ort-legal-regulatory-guidance-third-edition-june-2023.pdf?rev=d32c7583049048bab3d64e122e23cce4&hash=B46645F9C018DF75BB5847BE29E14F00

Thomas, R. & Zikopoulos, P. (2020). The AI Ladder: Accelerate Your Journey to AI. O'Reilly.

Thompson, J. (2023). *Data for All*. Manning Publications.

Torra, V. (2017). *Data Privacy: Foundations, New Developments and the Big Data Challenge*. Springer.

Treder, M. (2019). *Becoming a data-driven Organisation Unlock the value of data*. Springer Vieweg. https://doi.org/10.1007/978-3-662-60304-8

Treder, M. (2020). *The Chief Data Officer Management Handbook: Set Up and Run an Organization's Data Supply Chain*. Apress.

Trujillo, G. et al., (2015). *Virtualizing Hadoop: How to Install, Deploy, and Optimize Hadoop in a Virtualized Architecture*. VMWare Press

UN Human Rights Office - OHCHR. (2020). *How to make economic reforms consistent with human rights obligations*. https://www.ohchr.org/sites/default/files/GuidePrinciples_EN.pdf

UNCTAD. (2021). *DIGITAL ECONOMY REPORT 2021. Cross-border data flows and development: For whom the data flow*. https://unctad.org/system/files/official-document/der2021_en.pdf

UNECE (2022) *Informal document GRVA-12-11/Rev.1 12th GRVA, 24-28 January 2022 Agenda item 5(c) 1/16 Access to in-vehicle data*. https://unece.org/sites/default/files/2022-01/GRVA-12-11r1e.pdf

Ursic, H. (2018). The Failure of Control Rights in the Big Data Era: Does a Holistic Approach Offer a Solution? In M. Bakhoum, B. Conde Gallego, M.-O. Mackenrodt, & G. Surblytė-Namavičienė (Eds.), *Proprietary Rights in Digital Data? Normative Perspectives and Principles of Civil Law* (pp. 55-83). Springer. https://doi.org/10.1007/978-3-662-57646-5_4

Ustaran, H. (2019). Chapter 12: International Data Transfers. En E. Ustaran, & H. Lovells (eds.), *European Data Protection Law and Practice, 2da ed*. IAPP.

Van Loenen, B. (2018). Chapter 3 Towards a User-Oriented Open Data Strategy, En B.V. Loenen, G. Vancauwenberghe and J. Crompvoets (eds), *Open Data Exposed* (pp. 33-53). Springer.

Venkataramanan, N. & Shriram, A. (2017). *Data Privacy Principles and Practice*. CRC Press.

Verhulst, S., Young, A., & Srinivasan, P. (2015). *AN INTRODUCTION TO DATA COLLABORATIVES CREATING PUBLIC VALUE BY EXCHANGING DATA*. https://datacollaboratives.org/static/files/data-collaboratives-intro.pdf

Verhulst, S. & Young A. (2019). The Potential and Practice of Data Collaboratives for Migration. En A. Salah, A. Pentland, B. Lepri & E. Letouzé (eds.). *Guide to Mobile Data Analytics in Refugee Scenarios: The 'Data for Refugees Challenge' Study* (465-476). Springer Nature.

Viegas, V., et al. (2022) IT Security Controls: A Guide to Corporate Standards and Frameworks. Apress.

WEF. (2018). *Data Policy in the Fourth Industrial Revolution: Insights on personal data*. https://www3.weforum.org/docs/WEF_Data_Policy_in_the_Fourth_Industrial_Revolution_2020.pdf

WEF. (2019a). *APPA – Authorized Public Purpose Access: Building Trust into Data Flows for Well-being and Innovation*. http://www3.weforum.org/docs/WEF_APPA_Authorized_Public_Purpose_Access.pdf

WEF. (2019b). *Data Collaboration for the Common Good Enabling Trust and Innovation Through Public-Private Partnerships*. http://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf

WEF. (2019d). *The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value*. http://www3.weforum.org/docs/WEF_Next_Gen_Data_Sharing_Financial_Services.pdf

WEF. (2019c). *Exploring International Data Flow Governance Platform for Shaping the Future of Trade and Global Economic Interdependence*. http://www3.weforum.org/docs/WEF_Trade_Policy_Data_Flows_Report.pdf

WEF. (2019e). *Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data*. http://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf

WEF. (2020a). *A New Paradigm for Business of Data BRIEFING PAPER*. http://www3.weforum.org/docs/WEF_New_Paradigm_for_Business_of_Data_Report_2020.pdf

WEF. (2020b). *A Roadmap for Cross-Border Data Flows: Future-Proofing Readiness and Cooperation in the New Data Economy*. http://www3.weforum.org/docs/WEF_A_Roadmap_for_Cross_Border_Data_Flows_2020.pdf

WEF. (2020c). *Shaping the Future of the Internet of Bodies: New challenges of technology governance*. http://www3.weforum.org/docs/WEF_IoB_briefing_paper_2020.pdf

WEF. (2021a). *Data-driven Economies: Foundations for Our Common Future WHITE PAPER*. http://www3.weforum.org/docs/WEF_WP_DCPI_2021.pdf

WEF. (2021b). *Good Data: Sharing Data and Fostering Public Trust and Willingness*. http://www3.weforum.org/docs/WEF_Good_Data_Sharing_Data_and_Fostering_2021.pdf

WEF. (2021c). *Resetting Data Governance: Authorized Public Purpose Access and Society Criteria for Implementation of APPA Principles*. http://www3.weforum.org/docs/WEF_Resetting_Data_Governance_2021.pdf

WEF. (2021d). *Rebuilding Trust and Governance: Towards Data Free Flow with Trust (DFFT)*. http://www3.weforum.org/docs/WEF_rebuilding_trust_and_Governance_2021.pdf

WEF. (2021e). *Developing a Responsible and Well-designed Governance Structure for Data Marketplaces BRIEFING PAPER (DFFT)*.

https://www3.weforum.org/docs/WEF_DCPI_Governance_Structure_Towards_Data_Exchanges_2021.pdf

WEF. (2021f). *Towards a Data Economy: An enabling framework*. https://www3.weforum.org/docs/WEF_Towards_a_Data_Economy_2021.pdf

WEF. (2021g). *Empowered Data Societies: A Human-Centric Approach to Data Relationships*. https://www3.weforum.org/docs/WEF_Empowered_Data_Societies_2021.pdf

WEF. (2022). *Advancing Digital Agency: The Power of Data Intermediaries*. https://www3.weforum.org/docs/WEF_Advancing_towards_Digital_Agency_2022.pdf

Weitzenboeck, E. et al. (2022). "The GDPR and unstructured data: is anonymization possible?", *International Data Privacy Law, Volume 12, Issue 3, August 2022* (pp. 184–206), https://doi.org/10.1093/idpl/ipac008

Wellerstein, A. (2021). *Restricted Data The History Of Nuclear Secrecy In The United States*. The University Of Chicago Press.

Werbach, K. (2018). *The Blockchain and the New Architecture of Trust*. The MIT Press.

West, D. M., & Allen, J. R. (2020). *Turning Point: Policymaking in the Era of Artificial Intelligence*. Brookings Institution Press.

Woodard, J. (2020). *Enterprise GIS: Concepts and Applications*. CRC Press.

World Bank Group. (2021) 'Data for Better Lives', World Development Report 2021. https://openknowledge.worldbank.org/bitstream/handle/10986/35218/9781464816000.pdf

WP29. (2007). Dictamen 4/2007 sobre el concepto de datos personales. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_es.pdf

WP29. (2014). Dictamen 05/2014 sobre técnicas de anonimización. https://www.aepd.es/documento/wp216-es.pdf

Yadron, D. et al. (2016). Inside the FBI's encryption battle with Apple, *The Guardian*.

https://www.theguardian.com/technology/2016/feb/17/inside-the-fbis-encryption-battle-with-apple

Yuming, L. (2019). *Data Rights Law 1.0 The Theoretical Basis*. Peter Lang. https://doi.org/10.3726/b15737

Yuming, L. (2020). *Data Rights Law 2.0 The System Construction*. Peter Lang. https://doi.org/10.3726/b17032

Yuming, L. (2021). *Data Rights Law 3.0 The Legislative Prospect*. Peter Lang. https://doi.org/10.3726/b1842