EUROPEAN UNION AGENCY
FOR CYBERSECURITY

INGENIERÍA DE LA PROTECCIÓN DE DATOS

De la teoría a la práctica

ENERO DE 2022

ACERCA DE ENISA

La Agencia de la Unión Europea para la Ciberseguridad (ENISA) es la agencia de la Unión cuyo objetivo es alcanzar un elevado nivel común de ciberseguridad en toda Europa. Creada en 2004 y reforzada mediante el Reglamento sobre la Ciberseguridad de la UE, la Agencia de la Unión Europea para la Ciberseguridad contribuye a la política de seguridad cibernética de la UE, mejora la fiabilidad de los productos, servicios y procesos de TIC mediante programas de certificación de la ciberseguridad, coopera con los Estados miembros y con los organismos de la UE y ayuda a Europa a prepararse para los desafíos del día de mañana en materia de ciberseguridad. A través del intercambio de conocimientos, el desarrollo de capacidades y las campañas de sensibilización, la Agencia coopera con sus partes interesadas clave para fortalecer la confianza en la economía conectada, para impulsar la resiliencia de la infraestructura de la Unión y, por último, para proteger digitalmente a la sociedad y a la ciudadanía de Europa. Para obtener más información sobre ENISA y su trabajo, puede consultar: www.enisa.europa.eu.

DATOS DE CONTACTO

Para ponerse en contacto con los autores, utilice la dirección de correo isdp@enisa.europa.eu.
Las consultas de los medios de comunicación acerca de este documento deben realizarse a través de press@enisa.europa.eu.

COLABORADORES

Claude Castelluccia (INRIA)
Giuseppe D'Acquisto (Garante per la Protezione dei Dati Personali)
Marit Hansen (ULD)
Cedric Lauradoux (INRIA)
Meiko Jensen (Universidad de Ciencias Aplicadas de Kiel)
Jacek Orzeł (Escuela de Economía SGH de Varsovia)
Prokopios Drogkaris (Agencia de la Unión Europea para la Ciberseguridad)

EDITORES

Prokopios Drogkaris (Agencia de la Unión Europea para la Ciberseguridad)
Monika Adamczyk (Agencia de la Unión Europea para la Ciberseguridad)

AGRADECIMIENTOS

Deseamos dar las gracias a nuestros compañeros del Comité Europeo de Protección de Datos (CEPD), Subgrupo de Tecnología, y a los compañeros del Supervisor Europeo de Protección de Datos (SEPD), Unidad de Tecnología y Privacidad, por revisar este informe y aportar sus valiosas observaciones.

También nos gustaría dar las gracias a Kim Wuyts, Veronica Jarnskjold Buer, Konstantinos Limniotis, Paolo Balboni, Stefan Schiffner, José M. del Alamo, Irene Kamara y a nuestra compañera de ENISA, Athena Bourka, por sus revisiones y sus valiosos comentarios.

AVISO LEGAL

MENCIÓN DE COPYRIGHT

ISBN 978-92-9204-556-2, DOI 10.2824/09079

TRADUCCIÓN AL CASTELLANO

La traducción al castellano ha sido realizada por ENISA y revisada por la Agencia Española de Protección de Datos.

En caso de discrepancia, la versión en inglés será considerada la fuente original y su interpretación la que prevalece. La versión original se puede encontrar en:

https://www.enisa.europa.eu/publications/data-protection-engineering

ÍNDICE

RESUMEN

La evolución de la tecnología ha traído consigo nuevas técnicas para compartir, tratar y almacenar datos. Con ello han surgido nuevos modelos de tratamiento de los datos (incluidos los datos personales), pero también nuevos problemas y amenazas. Entre algunos de los problemas relacionados con la evolución de la protección de la privacidad y los datos asociados a las tecnologías y aplicaciones emergentes se encuentran: la falta de control y transparencia, la posibilidad de que se reutilicen los datos, la inferencia y reidentificación de los datos, la elaboración de perfiles y la toma de decisiones automatizada.

La aplicación de los principios de protección de datos del RGPD en estos contextos resulta difícil, ya que no puede hacerse de la manera tradicional e «intuitiva». Las operaciones de tratamiento deben replantearse, a veces de un modo radical (en proporción a la radicalidad de las amenazas), posiblemente con la definición de nuevos agentes y responsabilidades, y con un papel destacado de la tecnología como elemento de garantía. Deben integrarse garantías en el tratamiento con medidas técnicas y organizativas. Desde el punto de vista técnico, el problema está en convertir estos principios en requisitos y especificaciones tangibles mediante la selección, aplicación y configuración de las medidas y prácticas técnicas y organizativas adecuadas.

La ingeniería de protección de datos puede percibirse como una parte de la protección de los datos desde el diseño y por defecto. Su finalidad es respaldar la selección, el despliegue y la configuración de medidas técnicas y organizativas adecuadas para satisfacer los principios específicos de la protección de datos. Sin lugar a duda, depende de la medida, del contexto y de la aplicación y, en última instancia, contribuye a la protección de los derechos y libertades de los interesados.

Con el presente informe se ha llevado a cabo un análisis más exhaustivo de la ingeniería de protección de datos con la intención de apoyar a los profesionales y las organizaciones en la aplicación práctica de los aspectos técnicos de la protección de datos desde el diseño y por defecto. En este sentido, este informe presenta las tecnologías (seguridad) y técnicas existentes y analiza los posibles puntos fuertes y la aplicabilidad en relación con el cumplimiento de los principios de protección de datos establecidos en el artículo 5 del RGPD.

Sobre la base del análisis facilitado en el informe, a continuación, se exponen las siguientes conclusiones y recomendaciones para las partes interesadas pertinentes:

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos) deben debatir y promover buenas prácticas en toda la UE en relación con las soluciones más avanzadas de las tecnologías y técnicas pertinentes. Las instituciones de la UE podrían promover estas buenas prácticas mediante documentos públicos pertinentes.

La comunidad investigadora debe seguir explorando el despliegue de técnicas y tecnologías (de seguridad) que puedan respaldar la aplicación práctica de los principios de protección de datos, con el apoyo de las instituciones de la UE en lo relativo a la orientación normativa y la financiación de la investigación.

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos) y la Comisión Europea deben promover el establecimiento de

regímenes de certificación pertinentes, con arreglo al artículo 43 del RGPD, para garantizar una adecuada ingeniería de la protección de datos.

1. INTRODUCCIÓN

Los avances tecnológicos de los últimos años han influido en la forma en la que se comparten y se tratan nuestros datos personales. La evolución de la tecnología ha traído consigo nuevas técnicas para compartir, tratar y almacenar datos. Con ello han surgido nuevos modelos de tratamiento de los datos (incluidos los datos personales), pero también nuevas amenazas y dificultades para el usuario final a la hora de comprender y controlar ese tratamiento. La continua presencia en línea de usuarios finales ha dado lugar a un aumento del tratamiento diario de grandes cantidades de datos personales. Pensemos, por ejemplo, en las compras en línea o cuando utilizamos una aplicación móvil para llegar a un lugar determinado o para ponernos en contacto con amigos y familiares. El ciclo de vida de los datos se ha ampliado y en él participan ahora muchos agentes, por lo que, en última instancia, los usuarios finales no pueden comprender y controlar plenamente quién, durante cuánto tiempo y con qué finalidad tiene acceso a sus datos personales.

La evolución de la tecnología ha aportado nuevas técnicas para compartir, tratar y almacenar datos que han traído consigo nuevos problemas y amenazas

En muchos casos, estas nuevas tecnologías se han introducido sin realizar previamente una evaluación de sus repercusiones en la privacidad y la protección de los datos. En este contexto, el tratamiento de datos personales se caracteriza a menudo por la ausencia de una finalidad predeterminada y por el descubrimiento de nuevas correlaciones entre los fenómenos observados, por ejemplo, en el caso de los macrodatos o el aprendizaje automático. Este *modus operandi* entra básicamente en conflicto con los principios de necesidad y limitación de la finalidad, tal y como se establecen en el RGPD. Otro ejemplo lo encontramos en las tecnologías de cadena de bloques y de registro descentralizado que ofrecen la oportunidad de sustituir las transacciones basadas en la intermediación, pero a expensas de que se produzca una pérdida sustancial del control que las personas tienen sobre sus datos, que en la cadena siguen estando visibles para todos los participantes en la misma, siempre que esté activa o puede que incluso sin que lo esté. Esto, dependiendo, por supuesto, de cada caso, contradice el principio del RGPD de minimización de los datos y constituye un importante obstáculo para el ejercicio del derecho de supresión de los interesados. Por último, los sistemas de inteligencia artificial podrían estar facultados para tomar decisiones con cierto grado de autonomía con el fin de alcanzar objetivos específicos, por ejemplo, en la calificación de solvencia en el ámbito financiero. Dicha autonomía podría entrar en conflicto con los requisitos previos de intervención humana sobre las máquinas y autodeterminación, ambos en el núcleo de la protección de los datos personales y en el RGPD.

Como se explica también en [1], entre algunos de los problemas relacionados con la evolución de la protección de la privacidad y los datos asociados a las tecnologías y aplicaciones emergentes se encuentran: la falta de control y transparencia, la reutilización incompatible de los datos, la inferencia y reidentificación de los datos, la elaboración de perfiles y la toma de decisiones automatizada. La aplicación de los principios de protección de datos del RGPD en estos contextos resulta difícil, ya que no puede hacerse de la manera tradicional e «intuitiva». Las operaciones de tratamiento deben replantearse y rediseñarse, a veces de forma radical (en proporción a la radicalidad de las amenazas y las vías de ataque), posiblemente con la definición de nuevos agentes y responsabilidades, y con un papel destacado de la tecnología como elemento de garantía. Las medidas técnicas y organizativas adecuadas, así como las garantías, deben considerarse en la fase más temprana posible e integrarse en el tratamiento. Este es el alcance del concepto de protección de datos desde el diseño, consagrado en el artículo 25 del RGPD.

1.1 PROTECCIÓN DE DATOS DESDE EL DISEÑO

La protección de datos desde el diseño es una obligación jurídica desde la entrada en vigor del RGPD en 2018. Sin embargo, el concepto surgió hace varios años en el contexto de la ingeniería de la privacidad[1]. En aquel momento recibía el nombre de «Privacidad desde el diseño» y adquirió gran relevancia, considerándose un componente esencial de la aplicación práctica de la protección de la privacidad y los datos personales. Actualmente se considera un concepto polifacético: en los documentos jurídicos, por una parte, se describe generalmente en términos muy amplios como un principio general; por otra parte, los investigadores e ingenieros suelen equipararlo con la utilización de tecnologías de protección del derecho a la privacidad (PET). Sin embargo, la privacidad desde el diseño no es solo una lista de principios ni puede reducirse a la aplicación de tecnologías específicas. De hecho, se trata de un proceso en el que intervienen diversos componentes tecnológicos y organizativos, que aplican los principios de privacidad y protección de los datos mediante el despliegue adecuado y oportuno de medidas técnicas y organizativas que también incluyen tecnologías PET.

La obligación que se describe en el artículo 25 es que los responsables del tratamiento deben diseñar e integrar la protección de los datos en el tratamiento de los datos personales, con una configuración adecuada por defecto o disponible de otro modo durante todo el ciclo de vida del tratamiento. Tras la adopción del RGPD, el CEPD ha publicado una serie de directrices [2] sobre protección de datos desde el diseño y por defecto, además de asesoramiento sobre su aplicación. La principal obligación es la aplicación de las medidas adecuadas y de las garantías necesarias para la aplicación efectiva de los principios de la protección de datos y, en consecuencia, de los derechos y libertades de los interesados desde el diseño y por defecto. Mediante los distintos ejemplos facilitados es evidente que el desarrollo y la integración adecuados y oportunos de medidas técnicas y organizativas en las actividades del tratamiento de los datos desempeñan un papel importante en la aplicación práctica de los diferentes principios de la protección de datos.

La finalidad de la ingeniería de protección de datos es respaldar la selección, el despliegue y la configuración de medidas técnicas y organizativas adecuadas para satisfacer los principios de la protección de datos

La ingeniería de estos principios se refiere, no solo a las decisiones tomadas respecto al diseño de la operación de tratamiento, sino también a la selección, el despliegue, la configuración y el mantenimiento de las medidas y técnicas tecnológicas adecuadas. Estas técnicas respaldarían el cumplimiento de los principios de la protección de datos y ofrecerían un nivel de protección adecuado al nivel de riesgo al que estén expuestos los datos personales. La ingeniería de protección de datos puede percibirse como una parte de la protección de los datos desde el diseño y por defecto. Su finalidad es respaldar la selección, el despliegue y la configuración de medidas técnicas y organizativas adecuadas para satisfacer los principios específicos de la protección de datos. Sin lugar a duda, depende de la medida, del contexto y de la aplicación y, en última instancia, contribuye a la protección de los derechos y libertades de los interesados.

1.2 FINALIDAD Y OBJETIVOS

La finalidad general del presente informe es llevar a cabo un análisis más exhaustivo de la ingeniería de protección de datos con la intención de apoyar a los profesionales y las organizaciones en la aplicación práctica de los aspectos técnicos de la protección de datos desde el diseño y por defecto. En este sentido, en este informe se pretende presentar las tecnologías y técnicas (de seguridad) existentes y analizar los posibles puntos fuertes y la aplicabilidad en relación con el cumplimiento de los principios de protección de los datos. Este trabajo se lleva a cabo en el contexto de las tareas de ENISA, en virtud del Reglamento sobre la Ciberseguridad[2], para respaldar a los Estados miembros en aspectos específicos de

[1] Lo presentó la Dra. Ann Cavoukian y también era una idea evidente, si bien no se mencionaba explícitamente en la Directiva 95/46/CE ni en la Directiva sobre la privacidad y las comunicaciones electrónicas. Véase también el documento [6] EDPS Opinion 5/2018 "Preliminary Opinion on privacy by design" del Supervisor Europeo de Protección de Datos (SEPD).
[2] Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) n.º 526/2013 («Reglamento sobre la Ciberseguridad») http://data.europa.eu/eli/reg/2019/881/oj

ciberseguridad de la política y la legislación de la Unión en materia de privacidad y protección de los datos. Este trabajo tiene por objeto servir como base para realizar un análisis más detallado y específico de las categorías de tecnologías y técnicas identificadas, demostrando al mismo tiempo su viabilidad práctica.

1.3 ESTRUCTURA DEL DOCUMENTO

En la sección 2 del documento se analiza la relación existente entre la ingeniería de la protección de datos y la protección de datos desde el diseño, la evaluación del impacto de la protección de datos y las tecnologías de protección del derecho a la privacidad para cumplir los principios generales de protección de datos. En la sección 3 se analiza la anonimización y dos de las técnicas de anonimización más destacadas, y también se hace referencia a anteriores trabajos de ENISA en el ámbito de la seudonimización. En la sección 4 se analizan algunas de las técnicas disponibles, aparte del cifrado, en los ámbitos del enmascaramiento de datos y la computación para la preservación de la privacidad, mientras que en la sección 5 se describen las tecnologías para el control de accesos, el almacenamiento y las comunicaciones que preservan la privacidad. En la sección 6 se presentan medidas técnicas en el ámbito más general de la transparencia, la capacidad de intervención y las herramientas de control de los usuarios, mientras que en la sección 7 se concluye el documento y se formulan recomendaciones para futuros trabajos en este campo.

2. INGENIERÍA DE LA PROTECCIÓN DE DATOS

2.1 DE LA PROTECCIÓN DE DATOS DESDE EL DISEÑO HASTA LA INGENIERÍA DE LA PROTECCIÓN DE DATOS

El concepto de ingeniería de la privacidad y la protección de datos ya se ha descrito en el pasado, bien a través de un conjunto de estrategias de ingeniería dirigidas al principio de privacidad desde el diseño, bien como un conjunto de objetivos de protección de datos.

La ingeniería de la privacidad y la protección de datos ya se ha descrito en el pasado, bien a través de un conjunto de estrategias de ingeniería dirigidas al principio de privacidad desde el diseño, bien como un conjunto de objetivos de protección de datos

Por ejemplo, en su informe de 2015 [3], ENISA analizó el concepto de privacidad desde el diseño desde la perspectiva de la ingeniería. Además del análisis del concepto, en el informe, utilizando trabajos pertinentes sobre este campo, se presentaban ocho estrategias de protección de la privacidad desde el diseño, tanto orientadas a los datos como a los procesos, destinadas a preservar determinados objetivos en materia de privacidad. Como un modo diferente de enfocar la ingeniería de la privacidad y la protección de datos, se propuso un marco compuesto por seis objetivos con el fin de identificar salvaguardias para los sistemas informáticos que tratan datos personales. Además de la típica tríada de seguridad formada por los conceptos de «confidencialidad», «integridad» y «disponibilidad», también se propusieron tres objetivos adicionales: «desvinculación», «transparencia» y «capacidad de intervención». Asimismo, se publicaron importantes trabajos en el ámbito de la ingeniería de la privacidad en [4] y [5], y en el proyecto de investigación PRIPARE[3], financiado por la UE. A pesar de los diferentes puntos de partida de cada trabajo, en todos ellos se presentaron propuestas para vincular los requisitos de protección de datos con requisitos técnicos, ya sea mediante una metodología, objetivos específicos o un conjunto de estrategias que debían respetarse.

En su Dictamen preliminar 5/2018 [6] sobre la privacidad desde el diseño, el Supervisor Europeo de Protección de Datos (SEPD) presentó una descripción detallada de las metodologías de ingeniería de la privacidad como medio para traducir los principios de privacidad desde el diseño y por defecto. En ese mismo dictamen preliminar, el SEPD incluyó ejemplos de metodologías para identificar los requisitos de privacidad y protección de los datos e integrarlos en los procesos de ingeniería de la privacidad con vistas a aplicar las garantías tecnológicas y organizativas adecuadas. Algunas de estas metodologías definen los objetivos de protección de datos directamente a partir de los principios de privacidad y protección de datos, como los del RGPD, o bien los derivan de objetivos operativos intermedios. Otras metodologías se rigen por la gestión de riesgos.

2.2 CONEXIÓN CON LA EIPD

La Evaluación de impacto relativa a la protección de datos (EIPD) es uno de los requisitos introducidos en el RGPD y también puede percibirse como parte del planteamiento de «protección desde el diseño y por defecto». Además del énfasis que estos principios ponen en la necesidad de incluir la ingeniería de los requisitos de protección de datos en las operaciones de tratamiento, este énfasis también es evidente en el artículo 35, apartado 7, letra d), del RGPD. El legislador menciona explícitamente «*las medidas previstas para afrontar los riesgos, incluidas garantías, medidas de seguridad y mecanismos...*», que claramente van más allá del despliegue tradicional de medidas técnicas y organizativas, y exige un análisis, una selección y un uso más detallados de las técnicas capaces de garantizar el nivel de protección requerido. Es interesante señalar que estas disposiciones también están vinculadas al nivel de riesgo del

[3] http://pripareproject.eu/ . La publicación correspondiente está disponible en https://www.slideshare.net/richard.claassens/pripare-methodologyhandbookfinalfeb242016

tratamiento de los datos personales (que de nuevo funciona como un umbral para la adopción de las medidas pertinentes). Este concepto también es evidente en algunos de los planteamientos propuestos por las autoridades de protección de datos sobre la evaluación del impacto, como la Evaluación de impacto sobre la privacidad (EIP)[4] y las plantillas y orientaciones pertinentes de la EIPD facilitadas por otras autoridades nacionales de protección de los datos[5].

Las tecnologías de protección del derecho a la privacidadprivacidad se pueden clasificar en función de las características de la tecnología empleada en relación con los

2.3 TECNOLOGÍAS DE PROTECCIÓN DEL DERECHO A LA PRIVACIDAD

Las tecnologías de protección del derecho a la privacidad (PET, por sus siglas en inglés de *Privacy Enhancing Technologies*) abarcan la gama más amplia de tecnologías diseñadas para respaldar la aplicación de los principios de protección de los datos en el plano sistémico y fundamental. Como se describe en [7], las PET son «*un sistema coherente de medidas de TIC que protege el derecho a la privacidad suprimiendo o reduciendo los datos personales o evitando el tratamiento innecesario o indeseado de datos personales, sin menoscabo de la funcionalidad del sistema de información*». Las PET, como soluciones técnicas, pueden considerarse como los pilares básicos para cumplir los principios de protección de datos y las obligaciones que impone el artículo 25 del RGPD sobre la protección de datos desde el diseño. Por lo tanto, también son parte de los elementos que constituyen los pilares básicos de la ingeniería de protección de datos.

Puesto que, en función del contexto, el alcance y la operación de tratamiento en sí, las PET pueden consistir en una sola herramienta técnica o en un despliegue completo, es evidente que no existe un planteamiento válido para todo y que es necesario hacer una clasificación más detallada de las diferentes PET. En esta dirección, ENISA ha presentado una metodología [8] para analizar la madurez de las PET y un marco para evaluarlas y valorarlas en el contexto de las herramientas de privacidad en línea y en las plataformas móviles. Como puso de relieve la Agencia Española de Protección de Datos en [9], existen diversas iniciativas para la clasificación de las PET, bien sobre la base de sus características técnicas o de los objetivos que persiguen (en relación con los principios de protección de los datos que pueden respaldar).

Por lo que se refiere a las herramientas y tecnologías específicas, otra clasificación puede basarse en las características de la tecnología utilizada en relación con los datos tratados. Concretamente, estas características pueden ser:

- **Preservación de la verdad:** el objetivo de la ingeniería de la privacidad es preservar la exactitud de los datos, reduciendo al mismo tiempo su capacidad de identificación. Este objetivo puede alcanzarse, por ejemplo, diluyendo la granularidad de los datos (p. ej., la edad en lugar de la fecha de nacimiento). De este modo, los datos siguen siendo exactos, pero de una manera «minimizada», adecuada para el fin en cuestión. Asimismo, el cifrado puede considerarse una técnica de preservación de la verdad, ya que el cifrado aplicado en dirección inversa restablece completamente los datos originales sin introducir ninguna incertidumbre en el proceso.
- **Preservación de la inteligibilidad:** los datos se conservan en un formato que «tiene sentido» para el responsable del tratamiento, sin revelar los atributos reales de los interesados. Por ejemplo, el truco de introducir un desfase en una fecha de hospitalización mantiene el formato de día/mes/año, pero rompe el vínculo con los datos reales de un paciente identificado. Además, la inyección de ruido es una técnica de preservación de la inteligibilidad, ya que no altera el aspecto de los datos, lo cual garantiza la confidencialidad de los datos reales.
- **Tecnología operable:** posibilidad de ejecutar operaciones matemáticas y lógicas (p. ej., una suma o una comparación) en los resultados de sus aplicaciones. Operatividad no implica necesariamente inteligibilidad, ya que (como se afirma más adelante en

[4] https://www.cnil.fr/en/privacy-impact-assessment-pia
[5] https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/vurdere-personvernkonsekvenser/vurdering-av-personvernkonsekvenser/

este informe) existen diversas técnicas de cifrado en las que los resultados (no inteligibles) son directamente operables mediante operaciones que se pueden ejecutar sin problemas en el dominio cifrado.

Además de estas características, también puede llevarse a cabo una clasificación adicional con respecto a los principios de protección de datos del RGPD que puede respaldar cada categoría, al menos en teoría. Intentar llevar a cabo esta taxonomía podría resultar muy útil a los responsables y encargados del tratamiento de datos, ya que ofrecería un modelo de referencia sobre los fines para los que puede servir cada herramienta o técnica, o como indicación de lo que ya se ha logrado con las herramientas y técnicas ya implantadas. Cabe señalar, sin embargo, que el análisis global debe realizarse siempre por operación de tratamiento e incluir también aspectos como la naturaleza, el alcance, el contexto y los fines del tratamiento, similares al concepto de la EIPD.

3. ANONIMIZACIÓN Y SEUDONIMIZACIÓN

La anonimización y la seudonimización son dos técnicas muy conocidas que se utilizan ampliamente para poner en práctica principios de la protección de datos, como la minimización de los datos. La seudonimización también se menciona explícitamente en el RGPD como una técnica que puede respaldar la protección de datos desde el diseño (artículo 25 del RGPD) y la seguridad del tratamiento de datos personales (artículo 32 del RGPD). Sin embargo, estas técnicas suelen confundirse a menudo cuando, de hecho, existe una diferencia importante entre ellas y en su aplicación en la práctica. Como ya señaló el Grupo de trabajo del artículo 29 [10] y de acuerdo con el considerando (26) del RGPD, por información anónima se entiende información que no guarda relación con una persona física identificada o identificable y, por lo tanto, los datos anónimos no se consideran datos personales. Por el contrario, de conformidad con el artículo 4, apartado 5, los datos seudonimizados que puedan (re)atribuirse a una persona física utilizando información adicional son datos personales y se les aplican los principios de protección de datos del RGPD. Un error frecuente es considerar que los datos seudonimizados son equivalentes a los datos anonimizados.

No puede aplicarse un sistema genérico de anonimización a todos los casos de uso y ofrecer una protección plena e ilimitada

En el ámbito de la seudonimización, ENISA ha publicado en los últimos años una serie de informes [11], [12] y [13] en los que se trata el concepto y el papel de la técnica en el marco del RGPD, diferentes técnicas y modelos de seudonimización y una serie de casos de uso en los que se demuestra su aplicabilidad en la práctica. A tal fin, la presente sección se centra en la anonimización de los datos, con el objetivo de explicar brevemente la k-anonimización y la privacidad diferencial como dos técnicas posibles para anonimizar los datos relacionales o tabulares.

Respecto a los datos no tabulares o secuenciales, puede que la anonimización no sea tan fácil o directa. Por ejemplo, en el caso de los datos de movilidad, los estudios pertinentes [14], [15] y [16] han demostrado que conocer tres o cuatro puntos espacio-temporales de una trayectoria bastaba para reidentificar, con una alta probabilidad, a una persona entre una población de varios millones. Las posibles soluciones pueden consistir en publicar únicamente estadísticas sobre diferentes trayectorias o en proponer o publicar datos sintéticos, es decir, trayectorias generadas artificialmente a partir de las características estadísticas de las trayectorias reales [17]. Los datos sintéticos se explican con más detalle en la sección 4.5.

3.1 ANONIMIZACIÓN

La anonimización de los datos es un problema de optimización entre dos parámetros contradictorios: la utilidad de los datos y la protección de la reidentificación. De hecho, la anonimización de los datos se consigue modificando los datos, ya sea mediante la introducción de ruido o mediante la generalización. Normalmente, para conseguir una protección sólida de la reidentificación es necesario modificar considerablemente el conjunto de datos y, por lo tanto, repercutir negativamente en su utilidad. La anonimización de los datos implica, por tanto, encontrar el mejor equilibrio entre estos dos parámetros, y ese equilibrio depende a menudo de la aplicación y del contexto (es decir, cómo se distribuye y se utiliza el conjunto de datos). Como también se menciona en [10], [18] y [19], no debemos dar por sentado que un sistema genérico de anonimización vaya a ser apto para todos los casos de uso y que dicho sistema tendrá capacidad para ofrecer una protección plena e ilimitada. Cada solución debe adaptarse en función del tipo de datos, la operación de tratamiento, el contexto y los posibles modelos de ataque. Este concepto debe considerarse aplicable a todas las técnicas y tecnologías descritas en el presente documento. Como se menciona en el dictamen del Grupo de trabajo del

artículo 29 [10], «*Una solución de anonimización eficaz impide a todos singularizar a una persona en un conjunto de datos, vincular dos registros en un conjunto de datos (o dos registros pertenecientes a conjuntos diferentes) e inferir cualquier tipo de información a partir de dicho conjunto*».

En las dos secciones siguientes se incluye una descripción general rápida de los dos métodos de anonimización más conocidos, a saber, la k-anonimización y la privacidad diferencial ε. En [10], [20] y [21] se puede consultar una descripción más detallada de los sistemas de anonimización existentes.

La k-anonimización se basa en la idea de que la combinación de conjuntos de datos con atributos similares permite ocultar información identificativa de cualquiera de las personas

3.2 k-ANONIMIZACIÓN

El modelo de k-anonimización se introdujo a principios de la década de 2000 y se basa en la idea de que la combinación de conjuntos de datos con atributos similares permite ocultar información de cualquiera de las personas que contribuyen a esos datos. Tal y como se explica en [22], se considera que un conjunto de datos ofrece protección mediante k-anonimización si la información de cada interesado contenida en el conjunto de datos no puede distinguirse de al menos k-1 interesados cuya información también figure en el conjunto de datos. El concepto clave consiste en abordar el riesgo de reidentificación de datos anonimizados mediante la vinculación con otros conjuntos de datos disponibles. Por ejemplo, en la Tabla 1 se presenta un conjunto de datos de ejemplo.

Tabla 1: Conjunto de datos inicial

Nombre	Sexo	Código postal	Año de nacimiento	Enfermedad diagnosticada
George S.	M	75016	1968	Depresión
Martin M.	M	75015	1970	Diabetes
Marie J.	F	69100	1945	Trastorno del ritmo cardíaco
Claire M.	F	69100	1950	Esclerosis múltiple
Amelia F.	F	75016	1968	Ninguna
Annes J.	F	75012	1964	Artritis reumatoide
Sophia C.	F	75013	1964	Hemopatía
Simon P.	M	75019	1977	Sarcoidosis
Michael J.	M	75018	1976	Linfoma

Para anonimizar los datos del cuadro A se pueden utilizar diversas técnicas, como la supresión o la generalización[6]. En este ejemplo, el atributo *Sexo* se mantuvo inalterado, ya que se consideró importante para el estudio de las enfermedades. Por otra parte, los atributos de *Código postal* de la dirección del usuario y *Año de nacimiento* se generalizaron conservando únicamente el código postal del departamento y utilizando intervalos de diez años, respectivamente. Al tratar de k-anonimizar los datos con un valor *k* de dos (2) y con respecto al cuasi-identificador formado por el triplete {Código postal, Año de nacimiento, Sexo}, el conjunto de datos inicial es transformado, ya que por cada triplete de valores hay al menos dos entradas de la tabla original que se corresponden con él, tal y como se presenta en el Tabla 2 siguiente.

[6] La generalización también puede lograrse mediante la supresión de un atributo (columna).

Tabla 2: datos k-anonimizados (con *k=2*)

Código postal	Año de nacimiento	Sexo	Enfermedad diagnosticada
75***	[1960-1970]	M	Depresión
			Diabetes
69***	[1940-1950]	F	Trastorno del ritmo cardíaco
			Esclerosis múltiple
75***	[1960-1970]	F	Ninguna
			Artritis reumatoide
			Hemopatía
75***	[1970-1980]	M	Sarcoidosis
			Linfoma

El método de k-anonimización presenta varias limitaciones. Por ejemplo, el criterio de k-anonimización no protege contra ataques de homogeneidad, en los que todos los registros agrupados en una clase de equivalencia tienen el mismo valor sensible o uno similar. Se han introducido varias ampliaciones del modelo de k-anonimización para abordar esta cuestión, como la l-diversidad, que garantiza que por cada valor cuasi-identificador correspondiente a los datos *k*, habrá al menos l valores representativos de los datos sensibles [23] y [24]. La Tabla 2 es 2-anónimo y 2-diverso, porque siempre hay al menos dos enfermedades diferentes en un grupo de personas con el mismo cuasi-identificador. Sin embargo, si Simon P. tuviera linfoma en lugar de sarcoidosis, la Tabla 2 seguiría siendo 2-anónima, pero dejaría de ser 2-diversa. En este caso, se podría inferir que Michael J., que pertenece al grupo definido por (75, [1970-1980], M), tiene linfoma, mientras que, antes, esa predicción solo se podía hacer con una probabilidad del 50 % (1/2). Otra deficiencia de la k-anonimización es que no compone, es decir, varios conjuntos de datos k-anonimizados de las mismas personas pueden combinarse para reidentificar a dichas personas [25]. Por lo tanto, es muy difícil garantizar a priori el riesgo de reidentificación, que podría depender del conocimiento del adversario.

La garantía de protección en la k-anonimización depende del valor de k. Intuitivamente, un valor k alto ofrece mejor protección que un valor más bajo, pero a costa de la utilidad de los datos. Para seleccionar un parámetro para una definición de privacidad es necesario conocer el vínculo entre el valor del parámetro y el riesgo de que se produzca un incidente de privacidad. Como hemos visto anteriormente, calcular cuantitativamente dicho riesgo y, por tanto, el valor k correspondiente, es muy difícil en la k-anonimización [26]. En el ámbito de la asistencia sanitaria, en el que los datos médicos se comparten con un pequeño grupo de personas (normalmente con fines de investigación), algunas veces se elige un valor *k* comprendido entre 5 y 15[27]. Sin embargo, esta elección es muy arbitraria y *ad hoc*.

La privacidad diferencial permite estudiar tendencias estadísticas más amplias en un conjunto de datos, pero protege los datos de las personas que participan en ese conjunto de datos

3.3 PRIVACIDAD DIFERENCIAL

Los algoritmos de privacidad diferencial (PD) [28] pueden ofrecer garantías de que, tras analizar un conjunto de datos de varias personas, el resultado del análisis no se verá afectado y seguirá siendo el mismo, aunque los datos de cualquier persona (hasta ε) no se hayan incluido en el conjunto de datos. Dicho de otro modo, la privacidad diferencial permite estudiar tendencias estadísticas más amplias en un conjunto de datos, pero protege los datos de las personas que participan en ese conjunto de datos. El aprendizaje de estas tendencias (es decir, inferencias que son generalizables a una población de interés más numerosa) es probablemente el objetivo final de cualquier publicación de datos en general. La privacidad diferencial no es una técnica de anonimización *per se*, sino un modelo en el que pueden desarrollarse técnicas de anonimización y que permite cuantificar el riesgo de reidentificación.

Para que un proceso sea diferencialmente privado, debe modificarse ligeramente, por lo general con cierta aleatoriedad o ruido. Ese ruido tiene que calibrarse con el valor ε y la sensibilidad, es decir, en qué medida contribuye una persona al resultado del proceso. Aunque todavía no existe un método riguroso para escoger el parámetro clave ε [29], la mayoría de los sistemas que aplican la PD eligen un valor ε cercano al 1. En cualquier caso, debe elegirse el más bajo posible que ofrezca una utilidad y privacidad aceptables. Por ejemplo, supongamos que una ciudad desea publicar el número de personas que padecen una enfermedad crónica concreta. La publicación puede hacerse en forma de histograma, en el que cada segmento corresponda al recuento de personas de un determinado distrito que sufren la enfermedad. Cada persona puede, a lo sumo, influir en uno de estos segmentos con el valor 1 (la persona padece o no padece la enfermedad y solo vive en un distrito). Por lo tanto, la sensibilidad es de 1. Sabemos que esta publicación puede ser ε-diferencialmente privada con tan solo añadir ruido de Laplace de la escala 1/ε a cada recuento de segmentos [30]. La eficacia de la protección está determinada en gran medida por ese factor ε, por lo que la selección de su valor óptimo debe hacerse en el contexto de los datos, el tamaño de la población de usuarios del conjunto de datos y el tratamiento que se esté llevando a cabo.

Los procedimientos de anonimización deben combinarse con un análisis de riesgos y beneficios que defina los niveles aceptables de riesgo y rendimiento

Cabe destacar que la anonimización basada en la privacidad diferencial puede ser de dos tipos: anonimización global o local. En el modelo global de la privacidad diferencial, los datos los recopila un agregador central que los transforma, normalmente añadiendo ruido, con un mecanismo de privacidad diferencial. Este modelo requiere una confianza plena en el agregador. Sin embargo, en el modelo local, los usuarios participantes aplican un mecanismo de privacidad diferencial a sus propios datos antes de enviárselos al agregador. Por lo tanto, no exige esa confianza plena en el agregador. Normalmente, el modelo local requiere añadir más ruido y, por lo tanto, reduce la precisión, aunque también pueden emplearse técnicas de agregación seguras para minimizar la degradación en la precisión [31].

Una de las principales ventajas de la privacidad diferencial es que la pérdida de privacidad puede cuantificarse, incluso si un conjunto de datos determinado se anonimiza varias veces para fines o entidades diferentes (decimos que la «privacidad diferencial compone»). A modo de ejemplo, si un mismo conjunto de datos se anonimiza dos veces (p. ej., mediante dos entidades diferentes), cada una con un valor de privacidad de ε, sigue siendo diferencialmente privado, pero con un parámetro de privacidad de 2 ε. Otra característica importante de la privacidad diferencial es que admite postratamiento. Dicho de otro modo: el resultado del tratamiento de datos diferencialmente privados a través de una transformación fija sigue siendo diferencialmente privado.

3.4 SELECCIÓN DEL SISTEMA DE ANONIMIZACIÓN

La anonimización de los datos es un proceso complejo que debe llevarse a cabo caso por caso. Las posibles soluciones dependen de numerosos parámetros que varían de una aplicación a otra, como el tipo de datos (temporales, secuenciales, tabulares, etc.), la sensibilidad de los datos o los niveles aceptables de riesgo y degradación del rendimiento. Todo procedimiento de anonimización debe combinarse con un análisis de riesgos y beneficios que defina los niveles aceptables de riesgo y rendimiento. Este análisis de riesgos servirá de orientación al responsable del tratamiento a la hora de seleccionar el modelo, el algoritmo y los parámetros que utilizar.

La solución de anonimización adoptada también debe depender del contexto, por ejemplo, cómo se distribuirá el conjunto de datos anonimizados. El modelo de «publicación y olvido», en el que los datos anonimizados se divulgan públicamente sin control, requiere una mayor protección que el «modelo enclave», en el que los datos anonimizados los conserva el responsable del tratamiento de los datos y solo los pueden consultar investigadores cualificados. Algunas veces, los métodos de k-anonimización y privacidad diferencial se perciben como contrapuestos y se considera que debe utilizarse uno en detrimento del otro. Sin embargo, son métodos bastante complementarios que se adaptan a diferentes aplicaciones, tal

y como se explica en [32]. La k-anonimización es fácil de entender y se adapta bien a los datos tabulares. Es el método más indicado para publicar datos anonimizados que puedan utilizarse para diferentes fines. Sin embargo, dado que es vulnerable a una serie de ataques y a que su seguridad depende de los conocimientos previos de los adversarios, no es aconsejable utilizarlo en un modelo de «publicación y olvido».

El modelo de privacidad diferencial ofrece una protección más sólida que la k-anonimización debido a la aleatoriedad añadida de que es independiente de los conocimientos del adversario. A diferencia de la k-anonimización, la privacidad diferencial no necesita una modelización de los ataques y es segura, independientemente de lo que sepa el agresor. Por lo tanto, se adapta mejor al modo de «publicación y olvido». Sin embargo, la privacidad diferencial no se adapta bien a los datos tabulares y es más adecuada para publicar información estadística agregada (recuento de consultas, valores medios, etc.) sobre un conjunto de datos. Además, el sistema de anonimización basado en privacidad diferencial suele requerir su adaptación al uso de los datos. Resulta difícil generar un conjunto de datos anonimizado mediante privacidad diferencial que ofrezca una protección sólida y una buena utilidad para distintos fines [33]. Además, la privacidad diferencial ofrece un mejor rendimiento con conjuntos de datos en los que el número de participantes es grande, pero cada contribución individual es bastante limitada.

4. ENMASCARAMIENTO DE DATOS Y COMPUTACIÓN PARA LA PRESERVACIÓN DE LA PRIVACIDAD

Enmascaramiento es un término genérico que se refiere a funciones que, una vez aplicadas a los datos, ocultan su valor real. Los ejemplos más destacados son el cifrado y el *hashing*; sin embargo, dado que el término es bastante general, también abarca otras técnicas, algunas de las cuales se explican en esta sección. La principal utilidad del enmascaramiento en lo relativo a los principios de la protección de datos es la integridad y la confidencialidad (seguridad) y, en función de la técnica o del contexto de la operación de tratamiento, también puede incluir la responsabilidad proactiva y la limitación de la finalidad.

4.1 CIFRADO HOMOMÓRFICO

El cifrado homomórfico es un elemento fundamental de muchas tecnologías de protección del derecho a la privacidad, como la computación segura multiparte, la agregación de datos privados, la seudonimización o el aprendizaje federado en machine-learning, por citar algunas. El cifrado homomórfico permite realizar cálculos en datos cifrados sin tener que descifrarlos primero. El caso típico de uso del cifrado homomórfico es cuando un interesado desea externalizar el tratamiento de sus datos personales sin revelarlos en texto común. Es evidente que estas funcionalidades son muy prácticas cuando el tratamiento lo realiza un tercero, como un proveedor de servicios en la nube.

El cifrado homomórfico permite realizar cálculos en datos cifrados sin tener que descifrarlos primero

Existen dos tipos de cifrado homomórfico: parcial y completo [34]. El cifrado homomórfico parcial (PHE, por sus siglas en inglés de *Partially Homomorphic Encryption*) es aquel en el que solo puede realizarse una única operación en el texto cifrado, por ejemplo, suma o multiplicación. Por otra parte, el cifrado homomórfico completo (FHE, por sus siglas en inglés de *Fully Homomorphic Encryption*) admite múltiples operaciones (actualmente suma y multiplicación), lo que permite realizar más cálculos con los datos cifrados. El cifrado homomórfico constituye actualmente un acto de equilibrio entre utilidad, protección y rendimiento. El FHE ofrece una buena protección y utilidad, pero un rendimiento deficiente. El PHE, por el contrario, ofrece un buen rendimiento y una buena protección, pero su utilidad es muy limitada. Sin embargo, hay una trampa: el rendimiento del FHE es bastante ineficaz, ya que las operaciones más sencillas pueden tardar segundos u horas, dependiendo de los parámetros de seguridad [35].

La elección del tipo de cifrado homomórfico depende del nivel de protección deseado y de la complejidad de los cálculos que se vayan a realizar con los datos cifrados. Si las operaciones son complejas, el sistema de cifrado resultará más caro. La complejidad de un cálculo no se mide de la forma habitualmente utilizada en informática (tiempo y memoria), sino por la diversidad de operaciones (suma y multiplicación) realizadas en los datos. Si el cálculo solo requiere suma (como en la suma de algunos valores), podrá utilizarse un cifrado homomórfico parcial. Si el cálculo requiere alguna suma y un número limitado de multiplicaciones, podrá utilizarse un cifrado homomórfico medio, que es similar al cifrado homomórfico parcial, pero con una limitación en cuanto al número de operaciones que se pueden realizar en lugar de en cuanto a los tipos de operaciones. Si el cálculo requiere muchas sumas y multiplicaciones, deberá utilizarse un cifrado homomórfico completo.

4.2 COMPUTACIÓN SEGURA MULTIPARTE

La computación multipartita segura intenta resolver los problemas de confianza mutua entre un conjunto de partes de modo que ninguna de ellas pueda ver los datos de las demás

El concepto de computación multipartita segura (SMPC, por sus siglas en inglés de *Secure Multi-Party Computation*) se refiere a una serie de protocolos criptográficos que se introdujo en 1986 y que intenta resolver problemas de confianza mutua entre un conjunto de partes distribuyendo un cálculo entre esas partes sin que ninguna de ellas pueda ver los datos de las demás. Los protocolos de computación bipartita segura pueden, por ejemplo, calcular funciones de los datos de entrada de dos partes sin revelar los datos de una parte a la otra. Entre las variaciones destacadas de la SMPC figuran el Consenso bizantino [36], en el que la computación se amplía a diversas partes, y las subastas [37], en las que los participantes pueden pujar en una subasta sin revelar su oferta. Esta última ya se utiliza como una aplicación real[7] en Dinamarca, donde los agricultores daneses fijan entre ellos los precios de la remolacha azucarera sin necesidad de un subastador central.

El ejemplo más destacado de SMPC es la tecnología de cadena de bloques, en la que un conjunto de partes, denominadas «mineros», tienen que decidir y acordar cuál va a ser el siguiente bloque que enlazar al registro de la cadena de bloques. Este problema puede dividirse en dos subtareas:

a) Determinar el bloque (o conjunto de bloques) válido que enlazar a la cadena de bloques, y
b) Consensuar con los demás «mineros» de la cadena de bloques que este nuevo bloque es el que debe enlazarse.

La tarea a) la puede realizar cada minero individualmente. Esta tarea consiste en encontrar un valor *hash* válido que cumpla una serie de requisitos mediante búsqueda por fuerza bruta (que, en realidad, es la parte que más energía consume de la tecnología de cadena de bloques), y todavía no implica a diversas partes. Una vez que un minero encuentra y anuncia este valor *hash* válido, la mayoría de los mineros deben llegar a un consenso para que sea ese *hash* (y su nuevo bloque de transacciones) el que se enlace a la cadena de bloques. Esta tarea es similar al protocolo de consenso bizantino de Lamport, ya que algunos mineros podrían ser falsos o podrían proponer un valor *hash* y un bloque diferente que también cumplieran todos los requisitos.

En general, existen protocolos de computación segura multiparte para cada función que puede calcularse entre un conjunto de partes. Dicho de otro modo: si existe una forma de que un conjunto de partes pueda calcular conjuntamente los resultados de la función (intercambiando algunos mensajes y calculando algunos resultados intermedios locales), siempre existirá un protocolo multiparte seguro que resuelva el problema con las garantías de seguridad necesarias. Desgraciadamente, en muchos casos, un protocolo de computación segura multiparte puede acabar siendo una aplicación muy compleja y puede requerir una inversión considerable en la comunicación por red. Por lo tanto, probablemente no sea adecuado para aplicaciones con requisitos en tiempo real que deban cumplirse rápidamente.

Dependiendo del protocolo exacto elegido, los protocolos SMPC respaldan los objetivos de confidencialidad de la protección de la privacidad (ya que no se revelan las entradas de las otras partes) y de integridad (ya que ni siquiera los atacantes internos o externos pueden cambiar fácilmente el resultado del protocolo). De este modo, la energía total se distribuye entre todas las partes implicadas, que pueden ser muchas entidades en aplicaciones reales como la cadena de bloques. De esta forma, no resulta posible que una parte individual pueda unilateralmente determinar e imponer una decisión sobre las demás partes .

Además, dado que el protocolo SMPC seleccionado en una aplicación concreta deben conocerlo todas las partes implicadas, este planteamiento fomenta la transparencia en cuanto

[7] https://partisia.com/better-market-solutions/mpc-goes-live/

al tipo de tratamiento que se aplica a los datos de entrada. En el lado negativo de este planteamiento, sin embargo, está el hecho de que no es nada fácil anular manualmente el resultado de una computación SMPC en caso de error. Si, por ejemplo, se escribe una dirección de correo electrónico en un bloque de la cadena de bloques y los mineros acuerdan cuál es su bloque anfitrión mediante el protocolo de consenso, se convierte en parte de la cadena de bloques para siempre. Eliminar esa dirección de correo electrónico de la cadena de bloques más adelante será prácticamente inviable, ya que requeriría que todos los mineros la eliminen localmente del bloque e ignoren el error que la eliminación provocaría en los valores *hash* de la cadena de bloques modificada, lo que sería una infracción directa del protocolo de cadena de bloques.

4.3 ENTORNOS DE EJECUCIÓN CONFIABLES

El cifrado es una potente herramienta para proteger los datos, pero inservible si se vulnera la seguridad del dispositivo que se utiliza para almacenar, cifrar o descifrar los datos. Si fuera el caso, el adversario podría acceder a los materiales de descifrado y a los datos en texto común. Un entorno de ejecución confiable (TEE, por sus siglas en inglés de *Trusted Execution Environment*) puede desempeñar un papel clave en la protección de los datos personales impidiendo el acceso no autorizado, las filtraciones de datos y el uso de programas maliciosos. Ofrece protección contra adversarios fuertes que logran acceder a los dispositivos, ya sea físicamente o a distancia. Con un TEE, el tratamiento de los datos tiene lugar internamente en el entorno. Por lo tanto, es teóricamente imposible obtener ningún dato.

Un entorno de ejecución confiable (TEE) es un entorno de tratamiento inviolable en el procesador principal de un dispositivo. Al funcionar en paralelo al sistema operativo y utilizar tanto hardware como software, los TEE están diseñados para ser más seguros que los entornos de tratamiento tradicionales. También recibe el nombre de entorno de ejecución de sistema operativo enriquecido (REE), en el que se ejecutan el SO y las aplicaciones del dispositivo. Garantiza la autenticidad del código ejecutado, la integridad de los estados de ejecución (p. ej., registros, memoria y entradas/salidas de la CPU), así como la confidencialidad de su código, sus datos y sus estados de ejecución. El TEE puede resistir ataques de software, además de los ataques físicos perpetrados en la memoria principal del sistema. A diferencia de los coprocesadores de hardware dedicados, el TEE puede gestionar fácilmente su contenido mediante la instalación o actualización de su código y sus datos.

La recuperación de información privada permite al usuario consultar una entrada de una base de datos sin revelar cuál es el elemento consultado

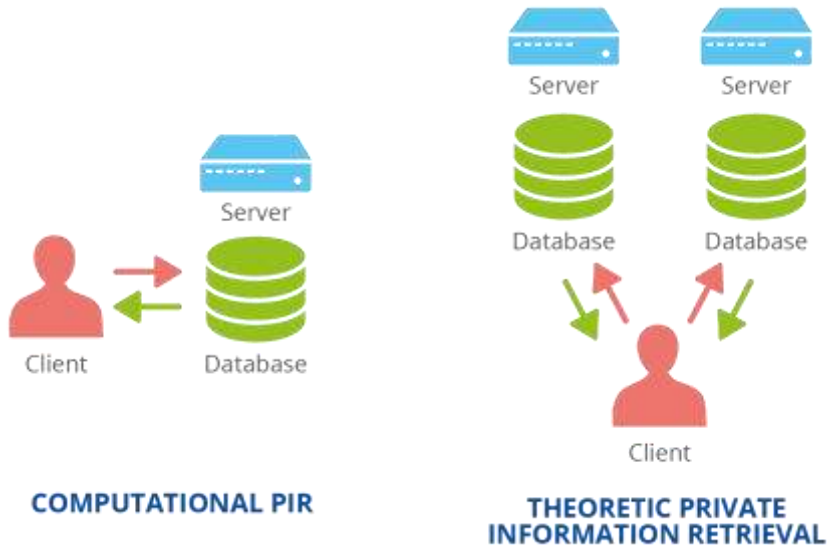
Los TEE se utilizan de forma generalizada en diversos dispositivos, como teléfonos inteligentes, tabletas y dispositivos de Internet de la Cosas (IoT). Los TEE también pueden desempeñar una importante función en la protección de los servidores. Pueden ejecutar funciones clave, como la agregación segura o el cifrado, para limitar el acceso a los datos sin procesar del servidor. También ofrecen la oportunidad de obtener cálculos verificables y aumentar la confianza. De hecho, los TEE permiten a los clientes certificar y verificar el código que se está ejecutando en un servidor determinado. En particular, una vez que el verificador conoce el código binario que se debe ejecutar en los enclaves protegidos, los TEE se pueden utilizar para verificar que un dispositivo está ejecutando el código correcto (integridad del código). Por ejemplo, en un entorno de aprendizaje colaborativo, los TEE y las certificaciones a distancia pueden ser especialmente útiles para que los clientes puedan verificar correctamente las funciones principales que se están ejecutando en el servidor, como la agregación segura o la aleatorización.

4.4 RECUPERACIÓN DE INFORMACIÓN PRIVADA

La recuperación de información privada (PIR, por sus siglas en inglés de *Private Information Retrieval*) es una técnica criptográfica que permite al usuario recuperar una entrada de una base de datos sin revelar al responsable de la custodia de los datos (p. ej., el propietario o administrador de la base de datos) el elemento que se ha recuperado [38]. Por lo tanto, los responsables del tratamiento pueden utilizarla como técnica de minimización de datos. Supongamos que una empresa desea que sus clientes puedan acceder a una base de datos.

En un entorno predeterminado, cada vez que un cliente accede a la base de datos, el custodio de los datos sabe a qué entrada ha accedido. Con el tiempo, el responsable del tratamiento podrá saber cuáles son las entradas de la base de datos que interesan a los clientes. Al implementar la técnica de recuperación de información privada, el responsable del tratamiento minimiza la cantidad de información revelada sobre los datos a los que se ha accedido, ya que la técnica PIR impide que el responsable del tratamiento sepa cuáles son las entradas consultadas.

Gráfico 1: Modelos de recuperación de información privada



Existen dos modelos principales de recuperación de información privada. El primero es el modelo de recuperación de información privada computacional, en el que la base de datos se aloja en un solo servidor. Se considera que este modelo ofrece una mejor protección, pero tiene limitaciones en cuanto a las conexiones que pueden establecerse con el servidor y la base de datos. El segundo es el modelo de recuperación de información privada teórica, en el que la base de datos está almacenada en varios servidores controlados por distintos propietarios. Este modelo admite una mayor complejidad de la comunicación, pero presupone que los servidores no coludirán ni intercambiarán información. Las referencias [39] y [40] incluyen información adicional sobre la PIR.

4.5 DATOS SINTÉTICOS

Los datos sintéticos se elaboran de modo que se asemejan de forma realista a los datos reales, pero en realidad no se refieren a ninguna persona concreta identificada o identificable

Los datos sintéticos son una nueva área del tratamiento de datos en la que los datos se elaboran de modo que se asemejan de forma realista a los datos reales (tanto personales como no personales), pero en realidad no se refieren a ninguna persona concreta identificada o identificable, ni a la medida real de un parámetro observable en el caso de los datos no personales. Por ejemplo, pueden ser datos totalmente simulados con el objetivo de probar servicios o aplicaciones informáticas. También pueden ser datos personales que se manipulan para limitar el potencial de reidentificación de las personas. El término «datos sintéticos» también se puede referir a la combinación de diversas fuentes de datos con el fin de disponer de mejores estimaciones de los parámetros de una población (una especie de enriquecimiento cruzado entre diferentes conjuntos de datos).

Al utilizar datos sintéticos, el responsable del tratamiento respeta la privacidad de las personas, ya que difieren de los datos reales, y la generación y el tratamiento de datos sintéticos no invade el ámbito personal de los interesados (sobre todo cuando los datos reales se refieren a características sensibles de las personas o a atributos raros que pueden ser difíciles de recuperar o que pueden tener un poder de identificación considerable). Sin embargo, esta

técnica también puede plantear problemas en términos de exactitud de los datos. Por lo tanto, los responsables del tratamiento siempre tendrán que reconciliar la tensión entre los diferentes principios de protección de datos, especialmente si el resultado del tratamiento implica consecuencias (es decir, consecuencias jurídicas o relacionadas con la salud) para los interesados. Los datos sintéticos deben adoptarse teniendo siempre en cuenta que se debe analizar, mediante un enfoque de ensayo y error, si su uso realmente genera estimaciones más precisas e imparciales a lo largo del tiempo. Desde esta perspectiva, los datos sintéticos son herramientas de ingeniería de la privacidad que pueden ofrecer datos granulares sin sacrificar la privacidad y la confidencialidad de los interesados.

Existen muchas alternativas prácticas para generar datos sintéticos. La opción más sencilla es extraer muestras de una distribución de probabilidad conocida. En este caso, el resultado no contiene datos originales (ni personales) y la reidentificación es poco probable, debido principalmente a la aleatoriedad. Otras opciones más complejas se basan en la combinación de datos reales y falsos (estos últimos obtenidos mediante muestreo de distribuciones multivariables conocidas, condicionadas por los datos reales observados). En este caso, sería posible revelar algunos datos personales y reidentificarlos debido a la presencia de valores reales dentro del conjunto de datos. Actualmente, la generación práctica de datos sintéticos, debido a la variedad de atributos implicados y al carácter diverso de las distribuciones de probabilidad, se basa tanto en el uso de rutinas clásicas de generación de números aleatorios como, cada vez más, en la aplicación de herramientas de inteligencia artificial y aprendizaje automático.

El uso de datos sintéticos presenta ventajas e inconvenientes, y los responsables del tratamiento deben ser conscientes de ambos. En lo referente a las ventajas, los datos sintéticos son datos que generan las máquinas y, como tales, su reproducción es sencilla y prácticamente exenta de costes. Los responsables del tratamiento no tienen que cargar con la tarea de la recopilación ni de la posible vulneración de los datos de los interesados. Además, los datos sintéticos también pueden ser válidos para situaciones en las que la recopilación de datos (personales) resulte muy difícil o incluso poco ética. Por ejemplo, en análisis comparativos en los que el objetivo es estudiar los efectos causales de una acción específica cuando realizar dicha acción puede que no sea una opción práctica. Podría ser una situación en la que se desee conocer el efecto que un nuevo tratamiento está teniendo en una patología o las consecuencias de la exposición a un factor de riesgo para la salud humana. En estas circunstancias, puede que no sea posible administrar el nuevo tratamiento a toda la población, o puede que no sea ético suspender el tratamiento anterior, y no es ético exponer deliberadamente a una persona a un factor de riesgo (p. ej., la contaminación) para comprobar sus efectos sobre su salud.

Los datos sintéticos pueden ayudar a los responsables del tratamiento a superar estas dificultades y poder llevar a cabo numerosos experimentos simulados. Además, los datos sintéticos (utilizados como una forma de anonimización) podrían beneficiarse de períodos de conservación más largos y potencialmente ilimitados.

Aunque hay mucho de cierto en ellos, es importante destacar que los datos sintéticos solo pueden imitar datos reales, reproduciendo propiedades específicas de un fenómeno, lo que significa que de ningún modo deben considerarse como medidas reales. Además, al tratarse de datos simulados, su calidad y exactitud dependen en gran medida de la calidad de los datos de entrada, que algunas veces proceden de fuentes dispares, y del modelo de ajuste de los datos. Los datos sintéticos también pueden reflejar los sesgos, tanto en las fuentes como en los modelos adoptados, de manera similar a los sesgos del aprendizaje automático. Por último, la generación de datos sintéticos no es una opción definitiva; requiere tiempo y esfuerzo. Aunque sean fáciles de crear, la producción de datos sintéticos debe controlarse, ya que su exactitud no está garantizada. Especialmente en situaciones complejas, la mejor forma de garantizar la calidad de su producción es comparar, a lo largo del tiempo, los resultados de los datos

sintéticos con los de datos auténticos etiquetados. Solo así se podrán reducir los riesgos de incoherencia. Pero, lo que es más importante, los datos sintéticos, debido a su naturaleza artificial, no son aptos para operaciones de tratamiento que impliquen a personas identificadas (como la elaboración de perfiles o cualquier decisión jurídicamente vinculante), sino más bien para análisis y predicciones generales.

Los ámbitos de aplicación de los datos sintéticos ya son numerosos y van en aumento, sobre todo si consideramos la necesidad de entrenar a los algoritmos de aprendizaje automático y los sistemas de inteligencia artificial con grandes volúmenes de datos en la fase de pruebas, antes de que pasen a formar parte de servicios o de un proceso productivo. Los datos sintéticos tabulares son datos numéricos que reflejan datos reales estructurados en tablas. El significado de los datos puede variar, desde datos relacionados con la salud hasta el comportamiento de los usuarios en la web o registros financieros. Un caso práctico de utilización de datos sintéticos tabulares sería en una empresa en la que los datos verdaderos no puedan distribuirse entre sus departamentos, filiales o socios debido a políticas internas o a limitaciones reglamentarias, mientras que su versión sintetizada podría utilizarse para hacer análisis predictivos.

5. ACCESO. COMUNICACIÓN Y ALMACENAMIENTO

5.1 CANALES DE COMUNICACIÓN

Los canales de comunicación seguros, como su propio nombre indica, son aquellos que permiten que se realice un intercambio seguro de datos entre dos o más partes que intervienen en una comunicación. Por lo general, están diseñados para mejorar la privacidad de las comunicaciones de manera que ningún tercero no autorizado pueda acceder al contenido y, en algunos casos, a los participantes o incluso a los metadatos de la comunicación en curso. El RGPD exige la seguridad de los datos personales tratados, incluso durante su transmisión, de conformidad con el considerando 49 y el artículo 32 del RGPD. La protección de la privacidad en el sector de las comunicaciones electrónicas y el tratamiento de metadatos, como los datos de tráfico, también está contemplada en la Directiva sobre la privacidad y las comunicaciones electrónicas[8].

Los canales de comunicación deben ir más allá de la prestación de seguridad como funcionalidad principal e incorporar características adicionales de protección del derecho a la privacidadprivacidad

Desde la perspectiva de la ingeniería de la protección de datos, los canales de comunicación deben ir más allá de la prestación de seguridad como funcionalidad principal e incorporar características adicionales de protección del derecho a la privacidad, como quién puede tener acceso al contenido de la comunicación, incluidos los proveedores, la ubicación y el acceso a las claves de cifrado, la ubicación y el tipo de proveedor, la información del usuario revelada, etc. A continuación, se describen dos tecnologías que van en esta dirección: el cifrado de extremo a extremo y el encaminamiento por *proxy*.

5.1.1 Cifrado de extremo a extremo

El cifrado de extremo a extremo (E2EE) es un método para cifrar los datos y mantenerlos cifrados en todo momento en la comunicación entre dos o más partes. Solo las partes implicadas en la comunicación tienen acceso a las claves de descifrado. La aplicación del cifrado de extremo a extremo es claramente una función fundamental de las aplicaciones de mensajería segura y ha ganado mucha fuerza en los últimos años, en los que una serie de servicios *Over-The-Top* (OTT) de uso generalizado, como las aplicaciones de mensajería, afirman aplicar el cifrado de extremo a extremo. La principal diferencia entre E2EE y el cifrado de enlaces o cifrado en tránsito es que, en el caso de estos dos últimos, el servidor puede acceder al contenido en función del lugar en el que se realice el cifrado o en el que se almacenen las claves de cifrado. En una situación típica E2EE, las claves de cifrado se almacenan en los dispositivos del usuario final y el servidor solo dispone de información sobre los metadatos de la comunicación (participantes en la comunicación, fecha/hora, etc.). Sin embargo, E2EE solo es fiable si no se vulnera la seguridad de uno de los puntos de conexión. Puede consultarse una descripción general de los protocolos de mensajes cifrados de extremo a extremo en [41] y [42].

A raíz de la sentencia C-311/18 (Schrems II)[9], que se refería a la transferencia de datos personales de un ciudadano de la UE a los Estados Unidos, el CEPD publicó sus recomendaciones [43] respecto a las medidas que complementan los instrumentos de transferencia para garantizar el cumplimiento del nivel de protección de los datos personales de

[8] 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 privacidadprivacidad en el sector de las comunicaciones electrónicas (Directiva sobre la privacidad y las comunicaciones electrónicas), revisada por la Directiva 2009/136/CE https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009L0136-20201221

[9] Sentencia de 16 de julio de 2020, *Schrems,* C-311/18, EU:C:2020:559, https://curia.europa.eu/juris/document/document.jsf?text=&docid=228677&pageIndex=0&doclang=en&mode=lst&dir=&occ=first&part=1&cid=40128973

la UE. En el Caso de uso 3 y en las condiciones que se mencionan en el mismo, el cifrado de extremo a extremo, combinado con el cifrado de la capa de transporte, se considera un medio para permitir transferencias de datos personales a países no pertenecientes a la UE en situaciones concretas.

5.1.2 Encaminamiento por *proxy* y de cebolla

Además del contenido de las comunicaciones que se ha descrito anteriormente, también se debe tener en cuenta el aspecto de los metadatos de la comunicación (datos que describen otros datos e incluyen información sobre quién, qué, dónde, cuándo, etc.) que, según la declaración del CEPD [44] respecto a la revisión de la Directiva sobre la privacidad y las comunicaciones electrónicas y la Propuesta de Reglamento sobre la privacidad y las comunicaciones electrónicas[10] «*pueden permitir extraer conclusiones muy precisas en cuanto a la vida privada de las personas, lo que implica un elevado riesgo para sus derechos y libertades*». En el artículo 4, apartado 3, letra c), de la propuesta se incluye una definición más formal de los metadatos de comunicaciones electrónicas.

Un posible modelo para proteger los metadatos es el uso de una red de encaminamiento de cebolla (por ejemplo, Tor[11]) que respalde la comunicación anónima a través de redes públicas. En el encaminamiento de cebolla, el tráfico de los usuarios se encamina a través de una serie de servidores de retransmisión [45] y cada servidor recibe los datos cifrados en capas sin saber quién es el remitente original ni el destinatario final. Esta información solo está disponible para el nodo de entrada y salida [46]. Sin embargo, Tor es vulnerable a los atacantes que pueden observar el tráfico que entra por los nodos de entrada y sale por los de salida y correlacionar los mensajes, tal y como se explica en [47].

El almacenamiento con protección de la privacidad protege la confidencialidad de los datos personales en reposo e informa a los responsables del tratamiento de datos en caso de infracción

5.2 ALMACENAMIENTO CON PROTECCIÓN DE LA PRIVACIDAD

El almacenamiento con protección de la privacidad tiene dos objetivos: proteger la confidencialidad de los datos personales en reposo e informar a los responsables del tratamiento de datos en caso de infracción. El cifrado es la técnica principal que se utiliza para proteger la confidencialidad de los datos frente a accesos no autorizados. En función de las limitaciones de los responsables del tratamiento, puede aplicarse en tres niveles diferentes: (i) nivel de almacenamiento, (ii) nivel de base de datos y (iii) nivel de aplicaciones.

Gráfico 2: Opciones de cifrado de la base de datos

El cifrado a nivel del sistema de archivos y de disco reduce los riesgos de que un intruso acceda físicamente al disco en el que se almacenan los datos. Este planteamiento tiene la

[10] 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) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017PC0010
[11] Proyecto Tor https://www.torproject.org/

ventaja de que es transparente para los usuarios de la base de datos, pero es un método de «todo o nada», ya que no es posible cifrar solo determinadas partes de la base de datos ni tampoco ofrece una granularidad mejor que la del cifrado a nivel de archivo. En esta solución solo hay una clave de cifrado de cuya gestión se encargan los administradores del sistema de la base de datos. La clave se encuentra en el servidor que aloja la base de datos y debe estar protegida mediante acceso con los máximos privilegios.

El cifrado también se puede aplicar a nivel de la base de datos. Este método ofrece más flexibilidad que la solución anterior y puede aplicarse con diferentes granularidades en tablas, entradas o campos. También se puede aplicar cuando algunos campos o atributos de datos son más delicados que otros (creencias políticas o religiosas, por ejemplo). Sin embargo, dado que las claves de cifrado deben almacenarse con la base de datos, un adversario que pueda conectarse al servidor en el que se aloja la base de datos podría utilizar una herramienta forense como Volatility[12] para recuperar las claves directamente de la memoria volátil.

En el cifrado a nivel de aplicación, todos los datos los cifra el cliente con sus propias claves de cifrado y después almacena las claves. Sin embargo, si distintos clientes van a compartir varias entradas de la base de datos, será necesario intercambiar las claves criptográficas, lo que puede poner en peligro su seguridad. Este problema se puede evitar si se utilizan sistemas de cifrado específicos, como el cifrado homomórfico. Las claves de cifrado ya no tienen que compartirse, puesto que se pueden realizar cálculos con los datos cifrados.

Por lo que se refiere a la notificación, los protectores de pila (o «*canaries*») son un conocido mecanismo de seguridad que se utiliza para detectar ataques de software y desbordamientos del búfer. El concepto de protector de pila puede trasladarse a la protección de datos personales. Inyectar un protector de pila en una base de datos implica introducir en ella valores falsos que se supone que no debe utilizar nadie. Por lo tanto, el acceso a esos valores debe vigilarse para detectar infracciones de la seguridad de los datos. También es importante señalar que no debe ser posible distinguir estos valores falsos de los reales. En una posible aplicación de este sistema habrá un servidor que almacene la base de datos y otro distinto que se encargue de tramitar las solicitudes a la base de datos. El servidor encargado de las solicitudes a la base de datos debe tener capacidad para identificar solicitudes de valores controlados, detectando así un posible ataque o infracción. Este modelo es especialmente adecuado para responsables del tratamiento que deseen utilizar sistemas de almacenamiento en la nube de terceros. Sin embargo, el responsable del tratamiento debe encontrar un buen equilibrio entre el número de entradas reales de la base de datos y el número de valores controlados (entradas falsas) y, en cualquier caso, estas técnicas no pueden considerarse una panacea para la pronta identificación de infracciones de la seguridad de los datos.

5.3 CONTROL DE ACCESOS, AUTORIZACIÓN Y AUTENTICACIÓN PARA LA PROTECCIÓN DEL DERECHO A LA PRIVACIDAD

La autenticación, la autorización y el control de accesos tienen por objeto impedir que se produzcan actividades no autorizadas o no deseadas mediante la aplicación de controles y restricciones sobre lo que pueden hacer los usuarios, sobre los recursos a los que pueden acceder y sobre las funciones que pueden realizar en los datos, incluida la visualización, modificación o copia no autorizadas. La autenticación confirma la identidad de un usuario que solicita el acceso a los datos, mientras que la autorización determina qué acciones puede llevar a cabo un usuario autenticado. El control de accesos se refiere a una técnica que garantiza que solo los usuarios autenticados puedan acceder a la información para la que están autorizados. Estos tres elementos están estrechamente relacionados y omitir siquiera uno de ellos puede debilitar el grado de protección de los datos, ya que los usuarios autorizados pueden acceder a ellos o realizar acciones no autorizadas.

[12] https://www.volatilityfoundation.org

En función del contexto y de las necesidades, algunos mecanismos de control de accesos parecen más adecuados que otros. Como también se explica en [48], en una situación en la que el tratamiento de los datos personales de los clientes con fines comerciales se realiza a través de un proveedor de almacenamiento en la nube en línea, el control de acceso discrecional (DAC, por sus siglas en inglés de *Discretionary Access Control*) puede emplearse para acceder a los datos de una solicitud de servicio específica, como un servicio de impresión y entrega. A través de DAC, cada usuario puede especificar sobre sus objetos los permisos que concede a otros usuarios o entidades externas. El DAC ofrece a los usuarios una flexibilidad avanzada para establecer las propiedades del control de accesos deseadas, pero, como dato negativo, depende en gran medida del conocimiento y la comprensión de los riesgos asociados por parte de los usuarios. Por otra parte, en un sistema de información hospitalaria, en el que se asigna a cada agente (doctores, personal de enfermería, personal administrativo) diferentes funciones con diferentes privilegios (p. ej., un médico puede acceder a los datos médicos de los pacientes), el método de control de accesos basado en funciones (RBAC, por sus siglas en inglés de *Role Based Access Control*) parece más indicado.

5.3.1 Credenciales basadas en atributos de protección del derecho a la privacidad

Las credenciales basadas en atributos (ABC, por sus siglas en inglés de *Attribute Based Credentials*) permiten autenticar una entidad mediante la autenticación selectiva de distintos atributos, sin revelar información adicional, que se suelen utilizar y que perfectamente podrían incluir datos personales. Por ejemplo, para que un proveedor de servicios permita el acceso a un servicio en línea, dicho proveedor debe verificar la edad de la persona que solicita el acceso. En lugar de preguntar su edad a la persona, el proveedor podría solicitar el valor de un atributo que indique si el interesado tiene más de 18 años o no. Para incluir incluso más características de protección de la privacidad, también se han propuesto credenciales basadas en atributos que protegen el derecho a la privacidad [49]. Esta técnica autentica al usuario mediante atributos, pero lo hace de una forma que minimiza los datos, ya que aporta atributos no vinculados (entre sí). El proyecto de investigación ABC4Trust [50] del programa H2020 ha llevado a cabo un importante trabajo de investigación en este ámbito y, desde entonces, se han realizado diversos despliegues, como el de la tarjeta IRMA[13]. Recientemente, el proyecto de investigación Decode[14] del programa H2020 ha puesto en marcha demostraciones piloto de explotación de credenciales basadas en atributos en situaciones de la vida real.

Las pruebas de conocimiento cero permiten a los usuarios demostrar que conocen una información secreta sin revelar nada sobre el secreto

5.3.2 Prueba de conocimiento cero

Las pruebas de conocimiento cero [51] son primitivas criptográficas que pueden utilizarse para imponer el cumplimiento de los principios de confidencialidad y minimización de los datos del RGPD. La idea fundamental de una prueba de conocimiento cero es permitir que un usuario (un interesado) demuestre a un servidor (responsable del tratamiento de datos) que conoce una información secreta sin revelar nada sobre ese secreto ([52]). Las pruebas de conocimiento cero se utilizan principalmente para aplicar sistemas de autenticación y en la norma ISO/IEC 9798-5[15] se proponen varios protocolos.

Las pruebas de conocimiento cero no solo imponen el cumplimiento de la confidencialidad, sino que también —en comparación con otros sistemas de autenticación, como el que utiliza el nombre de usuario y la contraseña— aplican el principio de minimización de los datos. En los sistemas de autenticación basados en contraseña, el usuario establece la contraseña y la comparte con un servidor.

[13] Aplicación IRMA https://irma.app/
[14] https://decodeproject.eu/
[15] ISO/IEC 9798-5:2009 Tecnología de la información — Técnicas de seguridad — Autenticación de entidades — Parte 5: Mecanismos que utilizan técnicas de conocimiento cero https://www.iso.org/standard/50456.html

Cuando el usuario desea autenticarse en el servidor, indica su contraseña, que se compara con la registrada en el servidor. Si un adversario desea hacerse pasar por el usuario, puede robar la contraseña al usuario o al servidor. En un sistema de autenticación de prueba de conocimiento cero, este riesgo se limita únicamente al usuario, ya que el servidor no conoce el secreto utilizado por el usuario para autenticarse. La técnica minimiza la cantidad de información que el servidor conoce sobre el usuario y, por lo tanto, reduce la superficie de ataque. Las pruebas de conocimiento cero son también un elemento fundamental de muchos protocolos de computación segura multiparte.

Las pruebas de conocimiento cero presentan dos variantes: interactivas [51] y no interactivas [53]. Las pruebas de conocimiento cero interactivas requieren varias comunicaciones entre el usuario y el servidor. Las pruebas de conocimiento cero no interactivas no requieren ninguna comunicación. Las pruebas de conocimiento cero no interactivas se utilizan con mucha frecuencia en aplicaciones de cadena de bloques.

6. TRANSPARENCIA, CAPACIDAD DE INTERVENCIÓN Y HERRAMIENTAS DE CONTROL DEL USUARIO

Ofrecer información precisa sobre la protección de datos no es una tarea fácil, ya que podría ser necesario simplificar la información y también podría generar malentendidos

Un elemento fundamental de cualquier concepto de protección de datos es la habilitación de las personas para ejercer por sí mismas sus derechos de protección de datos. Dicha habilitación implica tanto el acceso a la información sobre el tratamiento de los datos (transparencia) como la capacidad para influir en el tratamiento de su información personal en el terreno del responsable o encargado del tratamiento (capacidad de intervención). A este respecto, la comunidad investigadora en materia de privacidad aportó numerosos planteamientos y temas que pueden ayudar a aplicar estos derechos y servicios correlacionados en las instituciones dedicadas al tratamiento de datos. En este capítulo presentamos una selección de los más importantes.

El RGPD no solo exige transparencia en el tratamiento de los datos, sino también la necesidad de que las personas comprendan por qué se recogen sus datos personales y cómo se tratan, p. ej., si se transfieren a otras partes. Mientras que los diseñadores de sistemas o los responsables de la protección de datos tienen capacidad para comprender los sistemas y procesos del tratamiento de los datos, e incluso para exigir información detallada acerca de ellos, la mayoría de los usuarios no puede comprender lo que establece una especificación técnica o un documento jurídico, e incluso pueden sentirse abrumados cuando se les presenta información básica sobre el tratamiento de datos personales. (Artículos 13 y 14 del RGPD).

6.1 POLÍTICAS DE PRIVACIDAD

En el mundo en línea, un instrumento bien conocido para facilitar información a los usuarios es la política de privacidad (también denominada «declaración de protección de datos», «política de datos» o «aviso de privacidad», entre otras formas).

La primera recomendación del Grupo de trabajo sobre protección de datos del artículo 29 respecto al modo de informar a los usuarios en línea sobre cuestiones relativas a la protección de datos se deriva de sus recomendaciones publicadas en 2004 [54] y destaca la posibilidad de utilizar un planteamiento de varios niveles, empezando por la información esencial y ofreciendo más información, si así lo desea el usuario, mediante niveles adicionales. El planteamiento por niveles resulta especialmente útil para la presentación de la información en dispositivos móviles, donde sería complicado, si no imposible, leer un texto largo con toda la información sobre el tratamiento de los datos personales.

En 2017, el Grupo de trabajo sobre protección de datos del artículo 29 publicó unas directrices sobre transparencia [55] que hacían referencia a las obligaciones establecidas en el RGPD. En particular, el documento explica el significado del requisito del artículo 12, apartado 1, sección 1, del RGPD: «*El responsable del tratamiento tomará las medidas oportunas para facilitar al interesado toda información indicada en los artículos 13 y 14, así como cualquier comunicación con arreglo a los artículos 15 a 22 y 34 relativa al tratamiento, en forma concisa,*

transparente, inteligible y de fácil acceso, con un lenguaje claro y sencillo, en particular cualquier información dirigida específicamente a un niño».

Ofrecer información precisa sobre la protección de datos no es una tarea fácil, ya que podría ser necesario simplificar la información para que una persona media pueda entenderla y, por otro lado, porque la simplificación no debe dar lugar a malentendidos. En el caso de la información en lenguaje natural, se han propuesto varias métricas para medir la complejidad y la comprensibilidad, por ejemplo, de los periódicos o de las condiciones de los seguros[16]. Los responsables del tratamiento pueden utilizar estas métricas para controlar la comprensibilidad de su política de privacidad, aunque las autoridades de protección de datos no hayan recomendado todavía el uso de una métrica concreta. Para dirigirse específicamente a los niños, la Information Commissioner's Office ha publicado un código de buenas prácticas respecto a los servicios en línea [56].

Al diseñar la presentación de la política de privacidad deben tenerse en cuenta los dispositivos que pueden utilizar los usuarios, p. ej., el tamaño de la pantalla. Asimismo, la información debe ser accesible y estar diseñada de manera que se ajuste a las tecnologías asistenciales, de modo que no se excluya a las personas con discapacidad. En algunas situaciones, la información textual no es adecuada, p. ej., en llamadas telefónicas o, en algunos contextos, en hogares inteligentes o en vehículos conectados. Asimismo, deberá comprobarse toda comunicación establecida a través de la HCI relativa a la información facilitada a los usuarios respecto al tratamiento de los datos. Los controles y las pruebas de inteligibilidad (y la ausencia de «patrones oscuros») podrían requerir la intervención de los responsables de la protección de datos y de personas con conocimientos de usabilidad.

Cabe señalar que facilitar información no se limita a un único documento básico, como puede ser la política de privacidad, sino que también puede ofrecerse información personalizada *ad hoc* durante el uso real según esté diseñada en la interfaz hombre-ordenador (HCI). Este tipo de información puede influir en los usuarios a la hora de tomar una decisión sobre cuestiones importantes relativas a la protección de datos (p. ej., qué datos publicar en redes sociales) o sobre la concesión o la retirada de su consentimiento[17]. En el informe titulado «*Deceived by Design*», el Consejo Noruego de Protección de los Consumidores señaló que el diseño de la interfaz hombre-ordenador de muchas aplicaciones y servicios no es neutro, sino que emplea los denominados «patrones oscuros», que empujan a los usuarios a revelar más datos y a tomar decisiones precipitadas sobre los métodos de tratamiento de los datos [57].

Los iconos de privacidad pueden ser un buen método para complementar la información textual, o incluso para sustituirla, sobre el tratamiento de datos personales

6.2 ICONOS DE PRIVACIDAD

También podría lograrse una mejor comprensión si la información se transmitiera, no solo mediante un texto que requiera conocimientos y esfuerzo de lectura, sino también a través de símbolos gráficos (iconos). Los iconos son un método bien conocido para complementar la información textual o incluso a veces para sustituirla. El RGPD recoge esta posibilidad en el artículo 12, apartado 7. Hasta la fecha, no existe un conjunto de iconos normalizados que pueda combinarse con la información de los artículos 13 y 14, pero la comunidad investigadora ya ha presentado varias propuestas [58], [59], [60] y [61].

En el artículo 12, apartado 7, también se introduce el requisito de que la información que se presente mediante iconos deberá ser legible mecánicamente. La lectura mecánica podría facilitar la comprensión del significado general del icono y la obtención de más información

[16] P. ej., la fórmula LIX de legibilidad, elaborada por Carl Hugo Björnsson en 1971: LIX(texto) = Total de palabras/Frases + (Palabras largas x 100)/Total de palabras.
La fórmula CFP (índice de funcionalidad del contenido) sobre informatividad: CFR(texto) = Cantidad de etiquetas de palabras de contenido / Cantidad de etiquetas de palabras funcionales.
Fórmula Hix (Hohenheimer Verständlichkeitsindex), elaborada por la Universidad de Hohenheim, basada en la fórmula Amstad, 1. Neue Wiener Sachtext-Formel, SMOG-Index y LIX, https://klartext.uni-hohenheim.de/hix
[17] Puesto que por «consentimiento» se entiende «toda manifestación de voluntad libre, específica, informada e inequívoca por la que el interesado acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen» (Artículo 4, apartado 11, del RGPD), requiere información suficiente.

sobre el tratamiento de datos facilitada mediante el icono. Entre las ventajas de que una política de privacidad como tal (y no solo los iconos) sea legible mecánicamente se incluyen la opción de traducirla al idioma preferido (y entendido) por el usuario y la interpretación automática en la máquina del usuario, posiblemente cotejando la información de la política de privacidad con las preferencias o demandas configuradas en el dispositivo del usuario (véase la sección 6.4 sobre las señales de preferencia de privacidad).

Algunos responsables del tratamiento ya han introducido iconos diseñados por ellos mismos y los presentan en combinación con su política de privacidad. A falta de un conjunto de iconos normalizado, pueden ser de gran utilidad para los usuarios. Sin embargo, en cuanto se publiquen soluciones normalizadas para los iconos y la legibilidad mecánica, deberán utilizarse. Como ya se ha señalado, quedan varias cuestiones pendientes relativas a las políticas de privacidad, como la falta de definición de las prácticas recomendadas, la ausencia de normas relativas a la legibilidad mecánica o la creación de un conjunto de iconos definido. Además, todavía no se han reflejado bien situaciones en las que intervienen las nuevas tecnologías, como las tecnologías de sensores con interfaces de usuario restringidas, ausentes, complejas o dinámicas y, por lo tanto, sistemas de tratamiento de los datos difíciles de entender que pueden implicar a varios responsables del tratamiento o sistemas informáticos en constante cambio.

A través de las señales de preferencia en cuanto a la privacidad, los usuarios pueden expresar sus preferencias de una manera legible mecánicamente

6.3 POLÍTICAS AUTOVINCULANTES (STICKY POLICIES)

La finalidad de las políticas de privacidad (legibles mecánicamente) puede ampliarse para regular el propio tratamiento de datos. Se han presentado varias propuestas sobre información de las políticas vinculada a los elementos de datos, por ejemplo, complementando la metainformación con la intención de describir propiedades importantes (origen, destinatarios, finalidades, fecha del tratamiento, etc.) o de controlar técnicamente las operaciones de tratamiento de datos permitidas y no permitidas (acceso, transferencia, supresión, etc.). Una ventaja de las políticas autovinculantes es que combinan la organización técnica del tratamiento de datos y la transparencia. Si un sistema de gestión de la protección de datos regula todo el tratamiento de los datos en el lado del responsable del tratamiento, los cambios en el tratamiento invocan automáticamente cambios en la información de la política. Del mismo modo, las restricciones establecidas en la política pueden garantizarse automáticamente. Los derechos de acceso, las restricciones temporales o los desencadenantes basados en eventos pueden definirse en un lenguaje de políticas que se pueda ejecutar en el sistema de tratamiento de los datos y que pueda traducirse a lenguaje natural en la política de privacidad.

La propuesta más destacada a este respecto es el trabajo sobre «políticas autovinculantes» [62], en el que las políticas se «adhieren» a los datos y viajan «pegadas» a ellos en caso de que se transfieran. También se están utilizando métodos criptográficos para evitar que los destinatarios ignoren las políticas adjuntas. Sin embargo, no todos los tipos de políticas excluyen totalmente la posibilidad de que se haga un uso indebido de los datos personales. En la actualidad no existen soluciones normalizadas para estas políticas mecánicamente legibles que también controlen las operaciones de tratamiento de datos.

6.4 SEÑALES DE PREFERENCIA EN CUANTO A LA PRIVACIDAD

Aunque los responsables del tratamiento de datos elaboran políticas de privacidad para informar sobre el tratamiento de los datos y sobre aspectos de la protección de datos, dichas políticas no constituyen necesariamente una comunicación unidireccional. Desde la década de los noventa se han venido debatiendo diversas posibilidades para que los usuarios, como interesados, expresen sus preferencias de privacidad de una manera mecánicamente legible, tal y como se expone en [63]. Si bien todos los usuarios del mundo en línea o de otros ámbitos pueden expresar sus exigencias o deseos respecto a cómo deben tratarse sus datos personales, los responsables del tratamiento de datos normalmente no pueden (y no lo hacen) satisfacer demandas arbitrarias. En su lugar, se atienen a operaciones de tratamiento estandarizadas y predefinidas. La comunicación de señales de preferencia respecto a la

privacidad normalizadas y legibles mecánicamente desde el lado del usuario puede ser interpretada por servidores que se rijan por estas normas técnicas. Un primer ejemplo de este planteamiento, que se quedó obsoleto en 2018, es la especificación de la «Plataforma de Preferencias de Privacidad»[64], que permitiría expresar políticas de privacidad, mediante un lenguaje de especificación, para el servidor y la complementaria para el usuario [65].

A raíz de la idea de que el usuario pudiera expresar sus deseos o exigencias respecto a la privacidad en un formato mecánicamente legible, se desarrollaron diversos lenguajes y protocolos, principalmente en proyectos de investigación y prototipos.

Aunque estos lenguajes y protocolos suelen ser planteamientos exhaustivos y a menudo complejos, con numerosas funciones, para las aplicaciones prácticas parecía oportuno utilizar un enfoque más simplista. Un ejemplo destacado fue la norma «No rastrear» (DNT, del inglés *Do Not Track*)[18], en la que los usuarios podían expresar su deseo de que no se les rastrease a través de un campo de la cabecera HTTP. «DNT = 1» significa «Este usuario prefiere que no se le rastree en esta solicitud», mientras que «DNT = 0» significa «Este usuario prefiere permitir que se le rastree en esta solicitud». Una tercera posibilidad sería abstenerse de enviar un encabezamiento DNT[19] porque el usuario no tenga activada esta función. Una de las deficiencias que presentaba el estándar DNT era la falta de legislación de apoyo: si la cuestión de «aceptación del rastreo» o «denegación del rastreo» solo se puede expresar como una «preferencia» en lugar de como una exigencia clara y si únicamente los servidores «educados» responden correctamente, esta medida no contribuirá a lograr fiabilidad ni claridad para los usuarios ni para los proveedores de servicios.

Otro estándar relativo al rastreo es el denominado «Control global de la privacidad» (GPC, por sus siglas en inglés de *Global Privacy Control*)[20]. Este estándar permite a los usuarios enviar a un sitio web la señal de «no vender ni compartir» a través de su navegador para solicitar que sus datos no se vendan ni se compartan con ninguna otra parte distinta de aquella con la que el usuario pretende interactuar, salvo en los casos permitidos por la ley. Desde mediados de 2021, la señal de GPC está regulada en la Ley de privacidad del consumidor de California (CCPA) adaptada[21]. Los usuarios que deseen expresar una señal de «no vender ni compartir» pueden utilizar uno de los navegadores o extensiones admitidos. En el marco del régimen europeo de protección de datos, los proveedores de servicios no están actualmente obligados a aplicar protocolos específicos que interpreten las señales de preferencia de privacidad de los usuarios.

Los paneles de privacidad permiten ver a los usuarios el modo en el que un responsable del tratamiento de datos está tratando sus datos

En la era del internet de las cosas (IoT), el papel de las políticas legibles mecánicamente, así como de las señales de preferencia de privacidad, irá ganando importancia. Los proveedores de sitios web o servicios web deben respaldar las señales de preferencia de privacidad normalizadas y tener en cuenta y respetar los deseos expresados por los usuarios a la hora de decidir sobre el tratamiento de sus datos personales. No obstante, cabe señalar que el artículo 25, apartado 2, del RGPD exige la protección de los datos por defecto, sin necesidad de que los usuarios declaren explícitamente que no están de acuerdo con que sus datos personales se traten para elaborar perfiles, se compartan o se vendan: el responsable del tratamiento debe garantizar «que, por defecto, solo sean objeto de tratamiento los datos personales que sean necesarios para cada uno de los fines específicos del tratamiento. Esta obligación se aplicará a la cantidad de datos personales recogidos, a la extensión de su tratamiento, a su plazo de conservación y a su accesibilidad».

Si las preferencias de los usuarios, expresadas de una forma técnica y normalizada, no pueden cumplirse, por razones de transparencia, los responsables del tratamiento deben informarles (p.

[18] https://www.w3.org/2011/tracking-protection/
[19] https://www.w3.org/TR/tracking-dnt/
[20] https://globalprivacycontrol.org/
[21] https://oag.ca.gov/privacy/ccpa#collapse7b

ej., en su política de privacidad) de los motivos de dicho incumplimiento. Por ejemplo, en algunos países puede haber leyes que exijan períodos de conservación más largos de lo que cabría esperar. Además de las señales de preferencia de privacidad normalizadas, los navegadores de los usuarios pueden contener herramientas de privacidad a modo de complementos o tener una configuración específica respecto al rastreo, la minimización de datos de identificadores o los bloqueadores de secuencias de comandos. En el caso de que una configuración restrictiva pueda impedir el correcto funcionamiento de un sitio web o de un servicio web, los proveedores deben informar a los usuarios sobre las posibles limitaciones y ofrecer, como mínimo, funciones básicas a esos usuarios a quienes les preocupa su privacidad. En la práctica, esto significa que deben respetar las exigencias de privacidad de los usuarios, independientemente de que utilicen uno de los próximos estándares de señales de preferencia en cuanto a la privacidad u otras herramientas.

6.5 PANELES DE PRIVACIDAD

Los paneles de privacidad pueden utilizarse como un mecanismo para aumentar la transparencia y, posiblemente, la capacidad de intervención de los interesados. El objetivo de los paneles de privacidad es que los usuarios puedan ver el modo en el que un responsable del tratamiento de datos está tratando sus datos. A diferencia de la información facilitada en las políticas de privacidad, que a menudo incluyen descripciones bastante abstractas de las operaciones de tratamiento, los paneles de privacidad pueden servir para mostrar los datos personales reales a los que tiene acceso el responsable del tratamiento. Con ellos, los interesados pueden entender mejor qué datos personales que les conciernen se están tratando. También suelen informar de cuándo y a quién se comunican esos datos personales, p. ej., si se están transfiriendo datos personales a otras organizaciones o si (y posiblemente para qué fin) una persona ha accedido a elementos de datos personales concretos. Los interesados también pueden comprobar si sus datos personales, tal y como se muestran en el panel de privacidad, están desfasados, son erróneos, están incompletos o son excesivos, o si su revelación a terceros es probable o inesperada.

Los paneles de privacidad actuales los facilitan normalmente los responsables del tratamiento, quienes también deciden la cantidad de información que se presenta en ellos, cuántas explicaciones se ofrecen, p. ej., sobre los posibles riesgos, y qué opciones puede utilizar el usuario para adaptar los parámetros o modificar o suprimir sus datos personales. Los usuarios no pueden esperar que se les informe de todos los tipos de comunicaciones de sus datos, por ejemplo, si las autoridades policiales u otras autoridades públicas exigen (en su jurisdicción) acceso lícito a los datos personales del usuario, pero prohíben notificárselo. Asimismo, la información sobre las violaciones de datos podría excluirse de la presentación. En caso de que el tratamiento de los datos por parte del responsable incluya la elaboración de perfiles, el panel de privacidad debe aclarar qué datos personales se utilizan para el perfil del usuario y, en todo caso, qué información se va a obtener de los datos personales agregados.

Los responsables del tratamiento deben considerar el uso de paneles de privacidad que sean prácticos para los interesados y que cumplan los requisitos del RGPD. Como se ha explicado en la sección 5.3, el usuario debe autenticarse mediante un método fiable para evitar la comunicación de datos personales a personas no autorizadas. Si los paneles de privacidad del lado del usuario se llegasen a distribuir, los responsables del tratamiento deberían comprobar si estas herramientas podrían utilizarse como tecnologías de mejora de la transparencia.

En las siguientes secciones se describen dos tipos de paneles de privacidad: los paneles de privacidad del lado de los servicios y los del lado del usuario.

6.5.1 Paneles de privacidad del lado de los servicios

Google fue una de las primeras empresas en ofrecer un panel de privacidad propio[22], que actúa como punto de acceso central para que los titulares de cuentas puedan gestionar su configuración de privacidad. Para las empresas que utilizan publicidad personalizada, estos paneles de privacidad también pueden resultar útiles para explicar la elección de los anuncios («¿Por qué veo este anuncio?») y, también, para pedir al usuario que ajuste las categorías configuradas o inferidas de anuncios que supuestamente pueden interesarle. Este planteamiento centrado en el usuario es criticable, ya que solo ofrece una transparencia parcial. Por ejemplo, no suele explicarse detalladamente cómo se han generado los perfiles sobre los intereses del usuario y, además, el modelo comercial de la publicidad personalizada no cumple necesariamente el principio de minimización de los datos, sino que empuja a los usuarios a facilitar más información que puede utilizarse para la elaboración de perfiles.

Los paneles de privacidad también se consideran una funcionalidad de los portales de servicios para ciudadanos del sector público. Probablemente Estonia haya sido el primer país en ofrecer una herramienta que presenta un resumen de los datos personales y de quién ha accedido a ellos: el sistema RIHA[23] muestra los datos personales que se almacenan en las distintas bases de datos públicas y sistemas de información gubernamentales, indicando la finalidad y las personas que pueden acceder a ellos. Los ciudadanos estonios pueden ver qué funcionarios han accedido a sus datos personales. Esta información se recopila mediante los ficheros de registro de accesos. Los ciudadanos pueden controlar los accesos que no deben realizarse sin una razón justificada, tal como exige la legislación nacional [66].

El sector público alemán tiene previsto aplicar una funcionalidad similar con un panel de privacidad denominado «Datenschutzcockpit», tal y como se recoge en la «Onlinezugangsgesetz». El hecho de que un panel de privacidad se incluya en un portal para ciudadanos no significa necesariamente que toda la información personal deba almacenarse permanentemente en una base de datos central; también puede aplicarse como una vista centralizada de información descentralizada.

6.5.2 Paneles de privacidad del lado del usuario

Los paneles de privacidad del lado del usuario son aplicaciones cuyo control está en manos del usuario, p. ej., como herramienta en el dispositivo del usuario. Estas herramientas, denominadas «tecnologías de mejora de la transparencia», se han desarrollado en el marco de proyectos de investigación y tienen por objeto aumentar la transparencia sobre la divulgación de los datos personales de los usuarios a diferentes responsables del tratamiento, posiblemente con distintos seudónimos, además de ayudar al usuario a gestionar identidades. En [67] se incluye un resumen.

Este tipo de paneles de privacidad se beneficiarían de políticas legibles mecánicamente y de señales de preferencia respecto a la privacidad estandarizadas para que el resumen del tratamiento de los datos personales pueda basarse en información fiable facilitada por un responsable del tratamiento. Cabe esperar que la normalización de las políticas legibles mecánicamente y su distribución en la práctica conduzcan al desarrollo de paneles de privacidad del lado del usuario.

Los paneles de privacidad no solo pueden ser un medio para presentar una visión general de la información sobre el tratamiento de los datos personales, sino que también pueden ofrecer funciones que permitan cambiar la configuración de privacidad o ejercer sus derechos a los interesados.

[22] Cuenta de Google: myaccount.google.com
[23] Riigi infosüsteemi haldussüsteem: https://www.riha.ee/

6.6 GESTIÓN DE CONSENTIMIENTOS

La mayoría de los servicios web existentes en internet los explotan empresas que llevan a cabo el tratamiento sobre la base jurídica del consentimiento. Al registrarse para utilizar tales servicios, el usuario debe aceptar las condiciones de uso, expresando así su conocimiento y consentimiento al tratamiento de los datos en las condiciones descritas en esos documentos.

Lamentablemente, los servicios web suelen cambiar con frecuencia; se añaden nuevas funciones y se eliminan progresivamente las antiguas. Se incorporan nuevos socios comerciales que pueden actuar como encargados del tratamiento de datos. Cada vez que esto ocurra, será necesario validar si los documentos de las condiciones de uso en vigor cubren los cambios introducidos en las operaciones de tratamiento.

La gestión de consentimientos se refiere a la administración de los consentimientos de los usuarios respecto al tratamiento de sus datos personales

En estos casos, resulta inevitable que el responsable del tratamiento modifique sus condiciones de uso en consecuencia (y posiblemente su política de privacidad y otros documentos similares), p. ej., incluyendo a los nuevos socios comerciales en la lista de encargados del tratamiento de datos o añadiendo la nueva finalidad del tratamiento de dichos datos. Una vez hecho esto, los nuevos clientes deberán leer y aceptar las nuevas condiciones de uso, expresando así su consentimiento a estas normas de tratamiento de datos. Sin embargo, los clientes existentes que ya aceptaron las antiguas condiciones de uso no expresaron su consentimiento al cambio en las normas del tratamiento y no pueden retirar fácilmente su consentimiento anterior. Esta situación puede ser problemática.

Si distintos clientes han aceptado diferentes versiones de las condiciones de uso en momentos diferentes, su base jurídica respecto al tratamiento de los datos puede ser distinta. No puede considerarse que el usuario concede automáticamente su consentimiento para todos los cambios que se introduzcan en el tratamiento de los datos. Por lo tanto, es necesario pedir explícitamente a los clientes existentes que revisen las nuevas condiciones de uso y que vuelvan a dar su consentimiento. Según el tipo y la aplicación del servicio web de que se trate, esta medida puede ser problemática.

De cualquier modo, en una situación realista, es inevitable permitir que distintos clientes utilicen el mismo servicio con condiciones de uso diferentes y, por lo tanto, con otra cobertura del consentimiento. Por lo tanto, también resulta inevitable llevar un seguimiento de estas diferencias, es decir, registrar qué consentimiento para el tratamiento de datos se aplica a cada cliente. Esta práctica se suele considerar un aspecto fundamental de la gestión de los consentimientos.

6.7 OBTENCIÓN DEL CONSENTIMIENTO

En el caso de los servicios web que se utilizan a través de un navegador, el planteamiento *de facto* para obtener el consentimiento consiste en mostrar el texto de las condiciones de uso al cliente para que lo lea y añadir un botón en la parte inferior que declare «He leído y entendido estas condiciones de uso». Una vez que el usuario pulsa el botón, se considera que ha concedido explícitamente su consentimiento y esa información se registra en el perfil de usuario del cliente (o, en el peor de los casos, se registra implícitamente por el hecho de que se crea un perfil de usuario y se permite al usuario iniciar sesión).

Lamentablemente, este planteamiento presenta varios inconvenientes:

- Los usuarios suelen pulsar el botón sin leer ni comprender el documento («están cansados de dar consentimientos» [68] y [69]).
- Los usuarios con discapacidad no pueden leer ni comprender el documento.
- Los problemas de visualización del navegador pueden dificultar a los usuarios la lectura del documento.

- Los servicios que no se pueden utilizar a través de un navegador no pueden emplear este método.
- Puede que los servicios que se ejecutan en ordenadores integrados (p. ej., en vehículos o dispositivos IoT) no dispongan de una pantalla en la que mostrar el documento de las condiciones de uso.
- Puede que los servicios que se ejecutan en ordenadores integrados no tengan ningún botón ni otro mecanismo de respuesta con el que expresar el consentimiento.

No obstante, la obtención del consentimiento es obligatoria incluso en tales circunstancias, y la expresión del consentimiento debe concederse y documentarse de un modo válido para poder utilizarlo como base jurídica para el tratamiento de datos.

6.8 SISTEMAS DE GESTIÓN DE CONSENTIMIENTOS

Desde el punto de vista de su implementación, existen numerosos planteamientos para gestionar los consentimientos en situaciones dinámicas reales (véase, por ejemplo, [70], [71] y [69]). En el ámbito de aplicación de la asistencia sanitaria, en el que se solicita el consentimiento de los pacientes antes de proceder a una intervención médica, se han propuesto conceptos muy elaborados para la gestión del consentimiento. Por razones obvias, los médicos de estas instituciones tienen la necesidad apremiante de documentar dicha manifestación de consentimiento antes de cualquier intervención, ya que la ausencia de su consentimiento podría hacer que, por ejemplo, se considerase que una intervención quirúrgica ha infligido daños corporales.

En este sentido, se han elaborado numerosos planteamientos para gestionar el consentimiento de modo que se pueda documentar un consentimiento único para una operación o tratamiento médico, pero la mayoría se basan en un extenso texto legal con la firma manuscrita del paciente debajo. Los equivalentes digitales utilizan documentos electrónicos, testigos de autenticación como tarjetas de identidad personales y tecnologías como la cadena de bloques para almacenar de forma permanente versiones exactas de los documentos de consentimiento junto con la expresión del consentimiento de los usuarios [72]. En estos casos nos encontramos con una serie importante de requisitos de seguridad para recopilar y documentar estas expresiones de consentimiento electrónicas, p. ej., en lo relativo a la integridad y disponibilidad de los formularios de consentimiento.

A diferencia de lo que ocurre en el ámbito de la asistencia sanitaria, los servicios de internet tienen necesidades ligeramente diferentes en lo que respecta a la recogida de consentimientos:

a) Si se va a obtener el consentimiento relativo a un sistema o servicio que está en evolución permanente, los cambios en el servicio deberán documentarse y reflejarse en las condiciones de uso. Esto difiere de las intervenciones quirúrgicas del ámbito de la asistencia sanitaria, donde solo se recopila una vez.
b) Las firmas manuscritas rara vez se utilizan como expresión del consentimiento en los servicios informáticos. Por lo tanto, la atribución, autenticidad y validez de una expresión de consentimiento deben recogerse utilizando diferentes técnicas, como las firmas electrónicas reconocidas. [73]
c) En la mayoría de los casos, el consentimiento debe obtenerse antes de que exista el perfil del usuario y, por tanto, la contraseña o el testigo de autenticación. No obstante, la asociación entre el perfil del usuario (es decir, sus datos personales) y la expresión de su consentimiento (cuando hace clic en el botón de aceptación) debe documentarse y almacenarse de forma que no pueda manipularse.

Como aspecto positivo, la utilización de sistemas de gestión del consentimiento reduce el trabajo de gestión del responsable y de los encargados del tratamiento. Una vez que el sistema está en funcionamiento, la tarea de recabar el consentimiento se realiza de forma prácticamente automática, liberando a los costosos y escasos recursos humanos

especializados para que puedan dedicarse a otras tareas, como puede ser la decisión de si debe recabarse o no un nuevo consentimiento. Una segunda ventaja clara es la capacidad de integrar el sistema de gestión de consentimientos en otras herramientas de gestión, como los sistemas CRM, para apoyo de los procesos de auditoría y certificación, o para cuestiones jurídicas. En este caso, dependiendo de la magnitud de la integración, existe un enorme potencial para minimizar los esfuerzos necesarios, ya que la alternativa sería aplicar procedimientos de gestión manuales, los cuales requieren la intervención de costosos recursos humanos con experiencia.

Como aspecto negativo, los esfuerzos para aplicar e integrar dicho sistema de gestión de consentimientos pueden ser sustanciales. En función del grado de integración, poner en marcha la instalación de un sistema de este tipo puede requerir una cantidad considerable de recursos y puede que solo se amortice parcialmente más adelante. Cuanto mayor sea el intento de integración, más alto será el coste inicial, pero también mayor será el ahorro resultante a largo plazo. Este (frecuente) desequilibrio puede impedir que las empresas más pequeñas integren sistemas de este tipo.

Responder a una solicitud para ejercer el derecho de acceso puede resultar difícil en redes de tratamiento de datos grandes y complejas en las que intervengan numerosos encargados del tratamiento de datos

6.9 EJERCICIO DEL DERECHO DE ACCESO

Como se establece en el artículo 15 del RGPD, el responsable del tratamiento debe facilitar a todos los interesados acceso a los datos personales que les conciernen y que estén almacenados en los sistemas del responsable del tratamiento, así como a la medida en que los encargados del tratamiento participan en el tratamiento de dichos datos. La misma obligación se aplica a los encargados del tratamiento de datos. Sin embargo, la tarea de responder a dichas solicitudes de información sobre el derecho de acceso, especialmente en redes de tratamiento de datos grandes y complejas con multitud de encargados del tratamiento de datos (y posiblemente otros corresponsables del tratamiento de datos; véase el artículo 26 del RGPD) puede resultar muy difícil. Para solucionar esta cuestión, muchas empresas modernas aplican infraestructuras y servicios técnicos para automatizar la tramitación de estas solicitudes de acceso facilitando, por ejemplo, paneles de privacidad, tal y como se explica en la sección 6.5.

Gráfico 3: Solicitud para ejercer el derecho de acceso

Un servicio para el ejercicio del derecho de acceso facilitaría una interfaz con los interesados cuyos datos los esté tratando la organización en cuestión. Al activarlo, el servicio consultaría automáticamente los repositorios de datos de la organización, recopilaría todos los datos personales relativos a la persona que realiza la solicitud y entregaría al interesado el conjunto completo de datos recogidos. Lo ideal es que toda la tarea esté automatizada, de modo que no sea necesaria ninguna interacción manual (o se requiera una interacción mínima) por parte de la organización.

Disponer de un servicio automatizado para ejercer el derecho de acceso que forme parte de los sistemas de gestión internos o externos de una organización reduce considerablemente los

esfuerzos manuales necesarios en caso de que la cantidad de solicitudes de acceso sea excesiva. Mientras los empleados pueden verse fácilmente extralimitados durante los picos de demanda, la infraestructura técnica suele ser más fácil de ampliar. En función de la cantidad de solicitudes recibidas, un sistema automatizado puede suponer un ahorro significativo para la organización.

Al mismo tiempo, si a este servicio para ejercer el derecho de acceso se conectasen todos los almacenes de datos nuevos, colectores de datos o encargados del tratamiento adicionales que obtengan datos de una persona, también mejorarían las capacidades de gestión de los datos de la organización en su conjunto. Las solicitudes relativas a la ubicación de los datos, el reenvío de los mismos, los socios comerciales que participan en el tratamiento de los datos, etc., pueden responderse de forma bastante sencilla a partir de las infraestructuras de flujo de datos existentes, creadas y mantenidas para dicho servicio de ejercicio del derecho de acceso.

Desde el punto de vista negativo, implementar un servicio de este tipo requiere definir, elaborar y desplegar otros procesos además de los servicios funcionales básicos para el tratamiento de datos. Esto plantea una serie de problemas adicionales que deben tenerse en cuenta:

- **Autorización:** una persona solo está autorizada a ver e investigar sus propios datos personales, no los de otros interesados. Por lo tanto, deben existir algunos medios (técnicos y organizativos) de autenticación que permitan verificar la autorización de la persona que realiza la solicitud. Por supuesto, dicha autorización debe garantizar su validez, por lo que deben emplearse técnicas de seguridad de alto nivel para validar la identidad de las personas. Podrían requerirse sistemas como la validación de pasaportes, la autenticación de dos factores u otros medios similares.

- **Autorización delegada:** en algunos casos es posible delegar el derecho de acceso, p. ej., en el caso de menores de edad, tutores legales, abogados, etc. En tales casos, la autorización de la solicitud de ejercicio del derecho de acceso debe validarse, no solo verificando la identidad de la persona que realiza la solicitud, sino también los fundamentos jurídicos de la transferencia de la autorización. En función del tipo de delegación de derechos, esta tarea puede resultar arbitrariamente compleja (véase también más adelante).

- **Riesgo de violación de la seguridad de los datos:** revelar todo el conjunto de datos personales de un interesado a otro sin una autorización válida equivale a una violación grave de la seguridad de los datos, lo cual, a su vez, supone una violación del RGPD. Al mismo tiempo, puede haber un interés sustancial en dichos derechos de acceso por parte de otros agentes distintos del interesado, como pueden ser piratas informáticos, medios de comunicación, fuerzas y cuerpos de seguridad o familiares. Por lo tanto, el riesgo para la seguridad que supone la prestación de un servicio de este tipo no es despreciable.

- **Completitud:** como se puso de manifiesto en la sentencia judicial de Schrems, obtener todos los datos divulgados en respuesta a una solicitud de ejercicio del derecho de acceso no resulta fácil. En un estudio reciente se demostró que solo aproximadamente el 10 % de las empresas facilitaban el conjunto completo de datos de sus clientes cuando estos lo solicitaban en virtud del artículo 15 del RGPD [74]. Sin embargo, una respuesta incompleta a una solicitud de derecho de acceso constituye una infracción del artículo 15 del RGPD y, por lo tanto, dejaría sin efecto el ejercicio de dicho derecho de acceso. El principal problema a este respecto está en cómo identificar todos los datos que pertenecen a un determinado interesado en el enorme conjunto de almacenes de datos con los que suelen trabajar las grandes organizaciones responsables o encargadas del tratamiento de datos. Algunas veces, esta tarea consiste en consultar las bases de datos con el identificador de cliente del

interesado (si está disponible), pero también puede incluir examinar grandes cantidades de datos archivados, sistemas de ficheros, copias de seguridad, conjuntos de datos derivados u otro tipo de información a la que ya no se puede acceder de una forma directa y sencilla o que no está debidamente vinculada al identificador de cliente del interesado, en caso de que existiese dicho identificador.

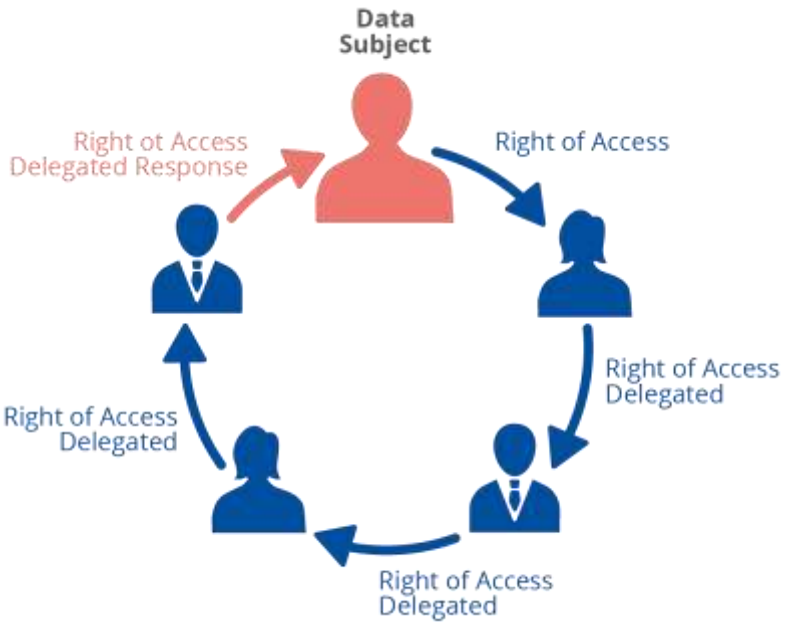
- **Exactitud:** al igual que la completitud, los datos comunicados al interesado deben ser exactos y, por lo tanto, no pueden contener abreviaturas, agregaciones, censura interna ni otros mecanismos que dificulten el acceso. Asimismo, debe mantenerse su integridad cuando se entreguen a la persona solicitante. Por lo tanto, para implementar un servicio para el ejercicio del derecho de acceso de este tipo se deben utilizar medios técnicos sólidos que garanticen la exactitud e integridad de los datos contenidos en la respuesta a la persona solicitante.

- **Volumen:** los perfiles personales de los interesados activos suelen aumentar de tamaño con la utilización de un servicio. Por lo tanto, la respuesta a una solicitud de ejercicio del derecho de acceso puede contener fácilmente enormes cantidades de datos. Esto plantea un reto técnico a la hora de entregar los datos a la persona solicitante por medios razonables. Por ejemplo, puede alcanzarse fácilmente el tamaño máximo permitido para su envío por correo electrónico, lo que hace que responder a una solicitud de ejercicio del derecho de acceso mediante un correo informativo resulte inviable. Imprimir los datos no solo sería problemático desde el punto de vista medioambiental, sino que tampoco cumpliría los requisitos comunes relativos a las respuestas a este tipo de solicitudes ni a la exigencia de portabilidad de los datos que se establece en el artículo 20 del RGPD. Las soluciones más utilizadas consisten en descargarse archivos de datos comprimidos a través de HTTP(S) o FTP(S), lo cual también implica algunos problemas técnicos en zonas con un ancho de banda deficiente.

6.9.1 Delegación de solicitudes de ejercicio de los derechos de acceso

Más allá de la facultad de delegar la autorización para ejecutar una solicitud de acceso, por ejemplo, a un custodio de datos, la delegación de una solicitud de derecho de acceso también puede incluir procesar una iteración en una serie completa de responsables y encargados del tratamiento de datos que participen en una actividad de tratamiento [75]. En estos casos, la delegación del derecho de acceso no solo se transfiere a una sola entidad específica, sino que básicamente se transmite junto con la propia solicitud a todos los subencargados implicados. Concretamente, una vez que el interesado solicita una información en virtud de sus derechos de acceso, el responsable o el encargado del tratamiento de los datos se pone en contacto con todos los subencargados implicados en ese tratamiento concreto y transmite a cada uno de ellos el derecho de acceso. Estos últimos, a su vez, también se ponen en contacto con sus subencargados y así sucesivamente, hasta que se haya contactado con toda la estructura de encargados del tratamiento (y corresponsables del tratamiento) implicados. A continuación, se responde a cada solicitud delegada con toda la información recibida de los subencargados del tratamiento más la información relativa a la organización a la que se solicitaron los datos. Por lo tanto, la solicitud presentada por el solicitante recibe una única respuesta obtenida tras recopilar, de forma recursiva, un conjunto completo de información procedente de todos los subencargados implicados.

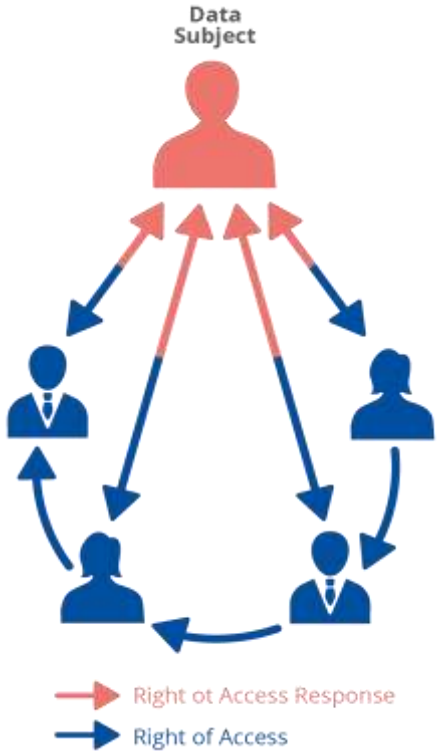
Gráfico 4: Delegación recursiva del derecho de acceso



Para el interesado, la ventaja de una infraestructura de este tipo es que una única solicitud para ejercer los derechos de acceso es suficiente para tener una idea completa de toda la actividad de tratamiento de los datos que realizan todos los subencargados del tratamiento. Para los encargados del tratamiento de datos, la ventaja de este tipo de infraestructura para la solicitud de datos es su composicionalidad: los detalles sobre la red exacta de subencargados, proveedores, proveedores de servicios, etc., pueden ocultarse o enmascararse fácilmente en la única respuesta enviada al anterior solicitante en la estructura de tratamiento. De este modo, la identidad exacta de los socios comerciales de una organización puede ocultarse a los encargados del tratamiento anteriores si se considera un secreto comercial. No obstante, los datos recogidos para la solicitud original de ejercicio del derecho de acceso siguen siendo completos y suficientes para satisfacer las necesidades del interesado.

El inconveniente evidente de este planteamiento es, de nuevo, los esfuerzos que requiere su implementación y su complejidad, ya que los servicios de derechos de acceso delegados deben implementarse y gestionarse. Cualquier solicitud de naturaleza recursiva puede requerir más recursos computacionales y plazos de ejecución más largos que la respuesta a una solicitud de derechos de acceso normal.

Gráfico 5: Delegación iterativa del derecho de acceso



Por otro lado, una organización encargada de la custodia de los datos también puede prestar por sí misma el servicio de recogida de datos para sus interesados. A diferencia del enfoque recursivo, en este caso la tarea de identificar individualmente y de exigir respuestas al derecho de acceso a todos los subencargados de la red de tratamiento la lleva a cabo de manera iterativa el custodio de los datos, en nombre y previa petición del interesado. Una vez finalizada la recogida, la respuesta agregada al derecho de acceso resultante se envía al interesado que la ha solicitado. En este caso, la ventaja de obtener una idea completa del tratamiento sigue siendo evidente para el interesado, mientras que los responsables y encargados del tratamiento pierden cierto control sobre lo que contiene exactamente dicha respuesta agregada a una solicitud de derechos de acceso.

6.10 EJERCICIO DE LOS DERECHOS DE SUPRESIÓN Y RECTIFICACIÓN

De forma similar al derecho de acceso, los demás derechos de los interesados, es decir, los derechos a la supresión, rectificación, bloqueo, limitación del tratamiento, etc., también pueden implementarse de forma parecida como servicios especiales o dedicados. En este caso, la infraestructura de los servicios para ejercer el derecho de acceso resulta muy útil, ya que permite identificar fácilmente todos los almacenes de datos afectados por una solicitud concreta. Además, permite enviar una notificación de que se ha iniciado una solicitud de derecho de supresión o rectificación a los encargados (o a los responsables) del tratamiento de dichos datos.

7. CONCLUSIONES

Los principios de protección de datos, establecidos en el artículo 5 del RGPD y elaborados en términos de medidas y garantías en el artículo 25, son los objetivos que deben alcanzarse al considerar el diseño, la implementación y el despliegue de una operación de tratamiento de datos. Desde el punto de vista técnico, el problema está en convertir estos principios en requisitos y especificaciones tangibles mediante la selección, implementación y configuración de medidas y prácticas técnicas y organizativas adecuadas durante todo el ciclo de vida del tratamiento de datos previsto. Sin embargo, la ingeniería de la protección de datos no es tan sencilla en la práctica. En función del nivel de riesgo, el contexto de la operación de tratamiento, los fines del tratamiento, los tipos, el alcance y el volumen de los datos personales, los medios y la magnitud del tratamiento, el estado de la técnica y el coste, la conversión en requisitos viables requiere un enfoque multidisciplinar. Además, el panorama tecnológico en evolución y las tecnologías emergentes también deben tenerse en cuenta a medida que surjan nuevos retos, como la falta de control y transparencia, la posible reutilización o desviación de la finalidad con el uso de datos, la inferencia y reidentificación de datos, la elaboración de perfiles y la toma de decisiones automatizada. La aplicación de los principios de protección de datos en estos contextos resulta difícil, ya que no puede hacerse de la manera tradicional e «intuitiva». Las garantías adecuadas, tanto técnicas como organizativas, deben integrarse en el tratamiento desde los primeros pasos, como dicta la obligación de protección de los datos desde el diseño, y el proceso de diseño y la correspondiente toma de decisiones también deben sustentarse en esta obligación.

En este informe se ha intentado ofrecer una breve descripción general de las tecnologías y técnicas (de seguridad) existentes que pueden ayudar a cumplir los principios de la protección de datos, además de explicar los posibles puntos fuertes y su posible aplicabilidad en diferentes operaciones de tratamiento. En el resto de esta sección se presentan las principales conclusiones a tal fin, junto con recomendaciones específicas para las partes interesadas pertinentes.

7.1 DEFINICIÓN DE LA TÉCNICA MÁS ADECUADA

Como ya se ha puesto de relieve en anteriores informes de ENISA, ya existen diversas tecnologías y técnicas, pero no resulta fácil para los responsables y encargados del tratamiento de datos saber cuál de ellas es aplicable y la más adecuada para cada operación de tratamiento y cada contexto. Y lo que es más importante, no está claro cómo debe integrarse cada técnica en la práctica de la operación de tratamiento con el fin de desarrollar realmente su potencial y respaldar el cumplimiento de los principios de la protección de datos.

La comunidad investigadora debe seguir explorando el despliegue de técnicas y tecnologías (de seguridad) que puedan respaldar la aplicación práctica de los principios de protección de datos, con el apoyo de las instituciones de la UE en lo relativo a la orientación normativa y la financiación de la investigación.

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos), la Comisión Europea y las instituciones pertinentes de la UE deben difundir las ventajas de estas tecnologías y técnicas, y ofrecer orientación sobre su aplicabilidad y despliegue.

Las iniciativas destinadas a apoyar a los ingenieros, como la Internet Privacy Engineering Network (IPEN)[24], deberían contar con el apoyo de profesionales, investigadores y el mundo académico.

7.2 DEFINICIÓN DEL ESTADO ACTUAL DE LA TÉCNICA

La correcta implementación e ingeniería de las tecnologías y técnicas analizadas depende en gran medida del estado de la técnica y de la manera en la que los responsables del tratamiento la conozcan o la tengan a su disposición. Aunque no todas las técnicas son igualmente eficaces, cada una de ellas puede presentar ciertos problemas o limitaciones de aplicación. Esto no solo afecta a la elección de la propia técnica, sino también al diseño general de la operación de tratamiento.

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos) deben debatir y promover buenas prácticas en toda la UE en relación con las soluciones más avanzadas de las tecnologías y técnicas pertinentes. Las instituciones de la UE podrían promover estas buenas prácticas mediante documentos públicos pertinentes.

7.3 DEMONSTRACIÓN DEL CUMPLIMIENTO Y OFRECIMIENTO DE GARANTÍAS

Además de orientación, los responsables y encargados del tratamiento de datos deben tener cierto grado de garantía respecto a la solidez y corrección de sus operaciones de tratamiento, cumpliendo al mismo tiempo su obligación reglamentaria de poder demostrar el nivel general de protección que ofrecen. En esta dirección, las disposiciones del artículo 42 del RGPD sobre mecanismos de certificación, sellos o marcas de protección de datos podrían convertirse en herramientas útiles, no solo para demostrar el cumplimiento (y la eficacia), sino también como guía a la hora de diseñar la protección de datos para las operaciones de tratamiento. Esto es aún más evidente en las tecnologías emergentes, como la inteligencia artificial, en las que las amenazas, los métodos tecnológicos, las implementaciones y los problemas de la protección de datos están evolucionando.

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos) y la Comisión Europea deben promover el establecimiento de regímenes de certificación pertinentes, con arreglo al artículo 42 del RGPD, para garantizar una adecuada ingeniería de la protección de datos.

Los reguladores (p. ej., las autoridades de protección de datos y el Comité Europeo de Protección de Datos) deben garantizar que los planteamientos reglamentarios, p. ej., en lo que respecta a las nuevas tecnologías y los sectores de aplicación, tengan en cuenta todas las entidades y funciones posibles desde el punto de vista de la protección de datos, sin dejar de ser tecnológicamente neutros.

[24] https://edps.europa.eu/data-protection/ipen-internet-privacy-engineering-network_en

8. REFERENCIAS

A.Will, M., & Ko, R. (2015). A guide to homomorphic encryption. En *The Cloud Security Ecosystem* (pág. 101127). Syngress.

Abul, O., Bonchi, F., & Nanni, M. (2008). Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. *IEEE 24th International Conference on Data Engineering (ICDE 08).*

Abul, O., Bonchi, F., & Nanni, M. (2010). Anonymization of moving objects databases by clustering and perturbation. *Information Systems, 35*(8), 849-910. doi:https://doi.org/10.1016/j.is.2010.05.003

Ács, G., & Castelluccia, C. (2011). I Have a DREAM! (DiffeRentially privatE smArt Metering). *International Workshop on Information Hiding (IH 2011).* doi:https://doi.org/10.1007/978-3-642-24178-9_9

Acs, G., & Castelluccia, C. (2014). A case study: privacy preserving release of spatio-temporal density in Paris. *20th ACM SIGKDD international conference on Knowledge discovery and data mining.* doi:https://doi.org/10.1145/2623330.2623361

Agencia Española de Protección de Datos (AEPD). (s.f.). *A Guide to Privacy by Design.* 2019. Obtenido de https://www.aepd.es/sites/default/files/2019-12/guia-privacidad-desde-diseno_en.pdf

ARTICLE 29 Data Protection Wokring Party. (2014). *Opinion 05/2014 on Anonymisation Techniques.* Obtenido de https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

ARTICLE 29 Data Protection Working Party. (2004). *Opinion 10/2004 on More Harmonised Information Provisions.* Obtenido de https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp100_en.pdf

ARTICLE 29 Data Protection Working Party. (2018). *Guidelines on transparency under Regulation 2016/679.* Obtenido de https://ec.europa.eu/newsroom/article29/items/622227

Asghar, M. R., Lee, T., Baig, M. M., Ullah, E., Russello, G., & Dobbie, G. (2017). A Review of Privacy and Consent Management in Healthcare: A Focus on Emerging Data Sources. *2017 IEEE 13th International Conference on e-Science (e-Science).*

Asonov, D. (2011). *Private Information Retrieval – An Overview and Current Trends.*

Balboni, P., & Francis, K. (2020). *Maastricht University Data Protection as a Corporate Social Responsibility (UM DPCSR) Research Project: UM DPCSR Icons Version 1.0.* Obtenido de https://www.maastrichtuniversity.nl/maastricht-university-data-protection-corporate-social-responsibility-um-dpcsr-research-project-um

Benchoufi, M., Porcher, .., & Ravaud, P. (2017). *Blockchain protocols in clinical trials: Transparency and traceability of consent.* F1000Research. Obtenido de https://pubmed.ncbi.nlm.nih.gov/29167732/

Blum, M., Feldman, P., & Micali, S. (1988). Non-interactive zero-knowledge and its applications. *Twentieth annual ACM symposium on Theory of computing (STOC 88).* doi:https://doi.org/10.1145/62212.62222

Bogetoft, P., Christensen, D. L., Damgård, I., Geisler, M., Jakobsen, T., Krøigaard, M., . . . Toft, T. (2009). Secure Multiparty Computation Goes Live. *Financial Cryptography and Data Security (FC 2009).*

Bonchi, F., Lakshmanan, L. V., & Wang, H. (2011). Trajectory anonymity in publishing personal mobility data. *SIGKDD Explor, 31*(1), 30-42.

Camenisch, J., Lehmann, A., Neven, G., & Rial, A. (2014). Privacy-Preserving Auditing for Attribute-Based Credentials. *European Symposium on Research in Computer Security (ESORICS 2014).* doi:https://doi.org/10.1007/978-3-319-11212-1_7

Chen, R., Acs, G., & Castelluccia, C. (2012). Differentially private sequential data publication via variable-length n-grams. *2012 ACM conference on Computer and communications security.* doi:https://doi.org/10.1145/2382196.2382263

Clifton, C., & Tassa, T. (2013). On syntactic anonymity and differential privacy. *IEEE 29th International Conference on Data Engineering Workshops (ICDEW).* IEEE.

Colesky, M., Hoepman, J. H., & Hillen, C. (2016). A Critical Analysis of Privacy Design Strategies. *2016 IEEE Security and Privacy Workshops (SPW).* doi:10.1109/SPW.2016.23

Datalisynet. (2015). *A guide to the anonymisation of personal data.*

Dwork, C. (2006). Differential Privacy. *International Colloquium on Automata, Languages, and Programming (ICALP 2006).* doi:https://doi.org/10.1007/11787006_1

Dwork, C., & Roth, A. (2014). *The Algorithmic Foundationsof Differential Privacy.* Foundations and Trends in Theoretical Computer Science. doi:10.1561/0400000042

Dwork, C., Kohli, N., & Mulligan, D. (2019). Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality, 9*(2). doi:https://doi.org/10.29012/jpc.689

E.Holtz, L., Nocun, K., & Hansen, M. (2010). Towards Displaying Privacy Information with Icons. *Privacy and Identity 2010: Privacy and Identity Management for Life.* doi:https://doi.org/10.1007/978-3-642-20769-3_27

EDPB. (2019). *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default.* Obtenido de https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en

Edwards, L., & Abel, W. (2014). *The Use of Privacy Icons and Standard Contract Terms for Generating Consumer Trust and Confidence in Digital Services.* Obtenido de http://zenodo.org/record/12506/files/CREATe-Working-Paper-2014-15.pdf

Emam, K. E., & Dankar, F. K. (s.f.). Protecting privacy using k-anonymity. *Journal of the American Medical Informatics Association, 15*(5), 627–637. doi: https://doi.org/10.1197/jamia.M2716

ENISA. (2015). *Privacy and Data Protection by Design.* Obtenido de https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

ENISA. (2015). *Privacy by design in big data.* Obtenido de https://www.enisa.europa.eu/publications/big-data-protection

ENISA. (2016). *PETs controls matrix - A systematic approach for assessing online and mobile privacy tools.* Obtenido de https://www.enisa.europa.eu/publications/pets-controls-matrix/pets-controls-matrix-a-systematic-approach-for-assessing-online-and-mobile-privacy-tools

ENISA. (2016). *Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies.* Obtenido de https://www.enisa.europa.eu/publications/pets

ENISA. (2017). *Security guidelines on the appropriate use of qualified electronic signatures.* Obtenido de https://www.enisa.europa.eu/publications/security-guidelines-on-the-appropriate-use-of-qualified-electronic-signatures

ENISA. (2019). *Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation.* Obtenido de https://www.enisa.europa.eu/publications/recommendations-on-shaping-technology-according-to-gdpr-provisions

ENISA. (2019). *Reinforcing trust and security in the area of electronic communications and online services: Sketching the notion of "state-of-the-art" for SMEs in security of personal data processing.* Obtenido de https://www.enisa.europa.eu/publications/reinforcing-trust-and-security-in-the-area-of-electronic-communications-and-online-services

ENISA. (2021). *Data Pseudonymisation: Advanced Techniques and Use Cases.* Obtenido de https://www.enisa.europa.eu/publications/data-pseudonymisation-advanced-techniques-and-use-cases

ENISA. (s.f.). *Pseudonymisation techniques and best practices.* Obtenido de https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices

Ermoshina, K., Musiani, F., & Halpin, H. (2016). End-to-End Encrypted Messaging Protocols: An Overview. *International Conference on Internet Science (INSCI).* Springer LNCS.

European Commission. (2017). *Factsheet: Access to Base Registries in Estonia.* Obtenido de https://joinup.ec.europa.eu/sites/default/files/inline-files/Estonia%20Factsheet%20Validated_0.pdf

European Data Protection Board. (2018). *Statement of the EDPB on the revision of the ePrivacy Regulation and its impact on the protection of individuals with regard to the privacy and confidentiality of their communications.* Obtenido de https://edpb.europa.eu/our-work-tools/our-documents/other/statement-edpb-revision-eprivacy-regulation-and-its-impact_en

European Data Protection Board. (2020). *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data.* Obtenido de https://edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-012020-measures-supplement-transfer_en

European Data Protection Supervisor. (2018). *Opinion 5/2018 Preliminary Opinion on privacy by design.* Obtenido de https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf

Fischer-Hübner, S., Angulo, J., Karegar, F., & Pulls, T. (2016). Transparency, Privacy and Trust–Technology for Tracking and Controlling My Data Disclosures: Does This Work? *Trust Management X: 10th IFIP WG 11.11 International Conference, IFIPTM 2016* (págs. 3-14). IFIP.

Forbrukerrådet (Norwegian Consumer Council). (2018). *Deceived by design.* Obtenido de https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/deceived-by-design/

Fung, B., Wang, K., Chen, R., & Yu, P. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys, 42*(4), 1-53.

Ganta, S. R., Kasiviswanathan, S. P., & Smith, A. (2008). Composition attacks and auxiliary information in data privacy. *14th ACM SIGKDD international conference on Knowledge discovery and data mining.* doi:doi.org/10.1145/1401890.1401926

Garfinkel, S. L. (2015). *NISTIR 8053 De-Identification of Personal Information.* Obtenido de https://nvlpubs.nist.gov/nistpubs/ir/2015/nist.ir.8053.pdf

Gasarch, W. (2004). A Survey on Private Information Retrieval. *Bulletin of the EATCS, 82*, 72-107.

Gentry, C., & Halevi, S. (2011). Implementing Gentry's Fully-Homomorphic Encryption Scheme. *Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2011).* doi:https://doi.org/10.1007/978-3-642-20465-4_9

Gilad, Y. (2019). Metadata-Private Communication for the 99%. *Communications of the ACM, 62*(9), 86-93.

Goldwasser, S., Micali, S., & Rackoff, C. (1985). The knowledge complexity of interactive proof-systems. *Seventeenth annual ACM symposium on Theory of Computing (STOC 85).* doi:https://doi.org/10.1145/22145.22178

Habib, H., Zou, Y., Yao, Y., Acquisti, A., Cranor, L., Reidenberg, J., . . . Schaub, F. (2021). Toggles, Dollar Signs, and Triangles: How to (In)Effectively Convey Privacy Choices with Icons and Link Texts. *2021 CHI Conference on Human Factors in Computing Systems.* doi:https://doi.org/10.1145/3411764.3445387

Hansen, M., Jensen, M., & Rost, M. (2015). Protection Goals for Privacy Engineering. *2015 IEEE Security and Privacy Workshops.* IEEE. doi:10.1109/SPW.2015.13

Herkenhöner, R., Meer, H. d., Jensen, M., & Pöhls, H. C. (2010). Towards Automated Processing of the Right of Access in Inter-organizational Web Service Compositions. *2010 6th World Congress on Services.*

Hils, M., Woods, D. W., & Böhme, R. (2020). Measuring the Emergence of Consent Management on the Web. *ACM Internet Measurement Conference.* doi:https://doi.org/10.1145/3419394.3423647

Hils, M., Woods, D. W., & Böhme, R. (2021). Privacy Preference Signals: Past, Present and Future. *Proceedings on Privacy Enhancing Technologies, 2021*(4), 249-269.

ICO. (2021). *Introduction to anonymisation.*

Information Commissioner's Office (ICO). (s.f.). *Age appropriate design: a code of practice for online services*. Obtenido de https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-a-code-of-practice-for-online-services/

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine Generals Problem. *SRI International*. Obtenido de https://lamport.azurewebsites.net/pubs/byz.pdf

Li, N., Li, T., & Venkatasubramanian, S. (2007). t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. *2007 IEEE 23rd International Conference on Data Engineering.* doi:10.1109/ICDE.2007.367856

Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2006). L-diversity: privacy beyond k-anonymity. *22nd International Conference on Data Engineering (ICDE'06).* doi:10.1109/ICDE.2006.1

Meyerson, A., & Williams, R. (2004). On the complexity of optimal K-anonymity. *23rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.*

Ostrovsky, R., & Skeith, W. E. (2007). A Survey of Single-Database Private Information Retrieval: Techniques and Applications. *PKC 2007: Public Key Cryptography.* Springer. doi:https://doi.org/10.1007/978-3-540-71677-8_26

P3P. (s.f.). *THE PLATFORM FOR PRIVACY PREFERENCES 1.1 (P3P1.1)*. Obtenido de https://www.w3.org/standards/history/P3P11

Quisquater, J.-J., Quisquater, M., Quisquater, M., Quisquater, M., Guillou, L., Guillou, M. A., . . . Guillou, S. (1990). How to Explain Zero-Knowledge Protocols to Your Children. *Advances in Cryptology (CRYPTO 89).* doi:https://doi.org/10.1007/0-387-34805-0_60

Rannenberg, K., Camenisch, J., & Sabouri, A. (2015). *Attribute-based Credentials for Trust.* Springer. doi:https://doi.org/10.1007/978-3-319-14439-9

Reed, M., Syverson, P., & Goldschlag, D. (1998). Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications, 16*(4), 482 - 494. doi:10.1109/49.668972

Reed, M., Syverson, P., & Goldschlag, D. (1998). Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications, 16*(4), 482-494.

S. Pearson, M. C.-M. (2011). Sticky Policies: An Approach for Managing Privacy across Multiple Parties. *Computer, 44*(9), 60-68. doi:10.1109/MC.2011.225

Santos, C., Nouwens, M., Toth, M., Bielova, N., & Roca, V. (2021). Consent Management Platforms Under the GDPR: Processors and/or Controllers? *Annual Privacy Forum 2021.*

Schermer, B. W., Custers, B., & Hof, S. v. (2014). The crisis of consent: how stronger legal protection may lead to weaker consent in data protection. *Ethics and Information Technology, 16*, 171-184.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(5), 557–570.

Tolsdorf, J., Fischer, M., & Iacono, L. L. (2021). A Case Study on the Implementation of the Right of Access in Privacy Dashboards. *Annual Privacy Forum 2021.* doi:https://doi.org/10.1007/978-3-030-76663-4_2

Unger, N., Dechand, S., Bonneau, J., Fahl, S., Perl, H., Goldberg, I., & Smith, M. (2015). SoK: Secure Messaging. *2015 IEEE Symposium on Security and Privacy.* IEEE.

W3C. (s.f.). *A P3P Preference Exchange Language 1.0 (APPEL1.0)*. Obtenido de https://www.w3.org/TR/2002/WD-P3P-preferences-20020415/

TP-01-22-045-EN-N

ACERCA DE ENISA

La Agencia de la Unión Europea para la Ciberseguridad (ENISA) es la agencia de la Unión cuyo objetivo es alcanzar un elevado nivel común de ciberseguridad en toda Europa. Creada en 2004 y reforzada mediante el Reglamento sobre la Ciberseguridad de la UE, la Agencia de la Unión Europea para la Ciberseguridad contribuye a la política de seguridad cibernética de la UE, mejora la fiabilidad de los productos, servicios y procesos de TIC mediante programas de certificación de la ciberseguridad, coopera con los Estados miembros y con los organismos de la UE y ayuda a Europa a prepararse para los desafíos del día de mañana en materia de ciberseguridad. A través del intercambio de conocimientos, el desarrollo de capacidades y las campañas de sensibilización, la Agencia coopera con sus partes interesadas clave para fortalecer la confianza en la economía conectada, para impulsar la resiliencia de la infraestructura de la Unión y, por último, para proteger digitalmente a la sociedad y a la ciudadanía de Europa. Para obtener más información sobre ENISA y su trabajo, puede consultar: www.enisa.europa.eu.

ENISA
European Union Agency for Cybersecurity

Athens Office
Agamemnonos 14, Chalandri 15231, Attiki, Greece

Heraklion Office
95 Nikolaou Plastira
700 13 Vassilika Vouton, Heraklion, Greece

enisa.europa.eu

ISBN: 978-92-9204-556-2
DOI: 10.2824/09079