III Insight Club DPO RGPD – LOPDGDD

La anonimización de datos: aclarando conceptos

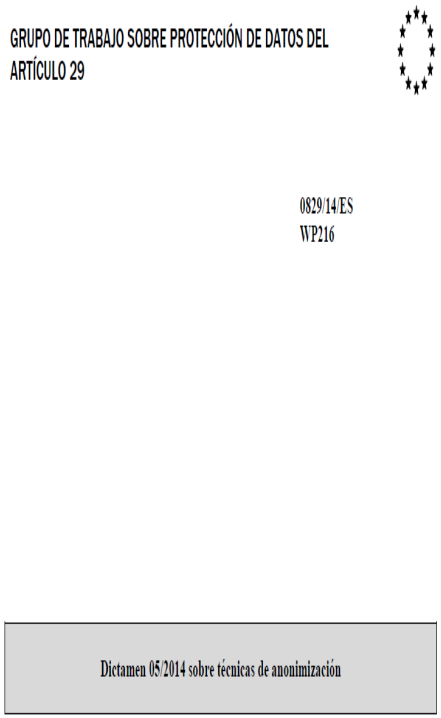
Francisco González-Calero
Lead Advisor at Govertis.
@FGonzalezCalero

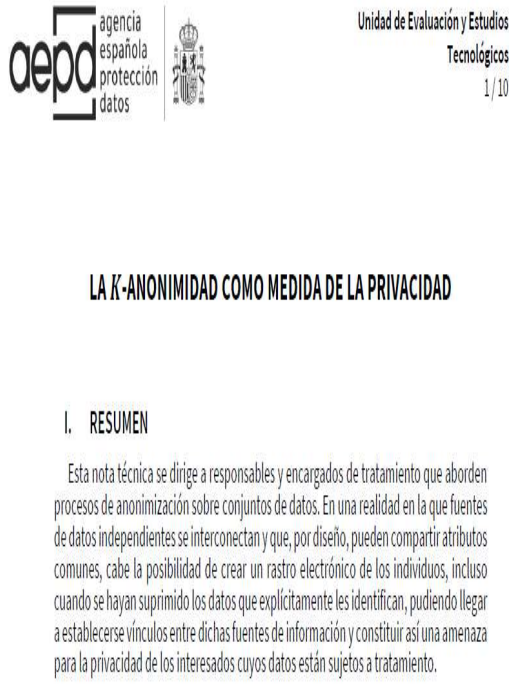
GOVERTIS Advisory Services

AEC
ASOCIACIÓN ESPAÑOLA PARA LA CALIDAD

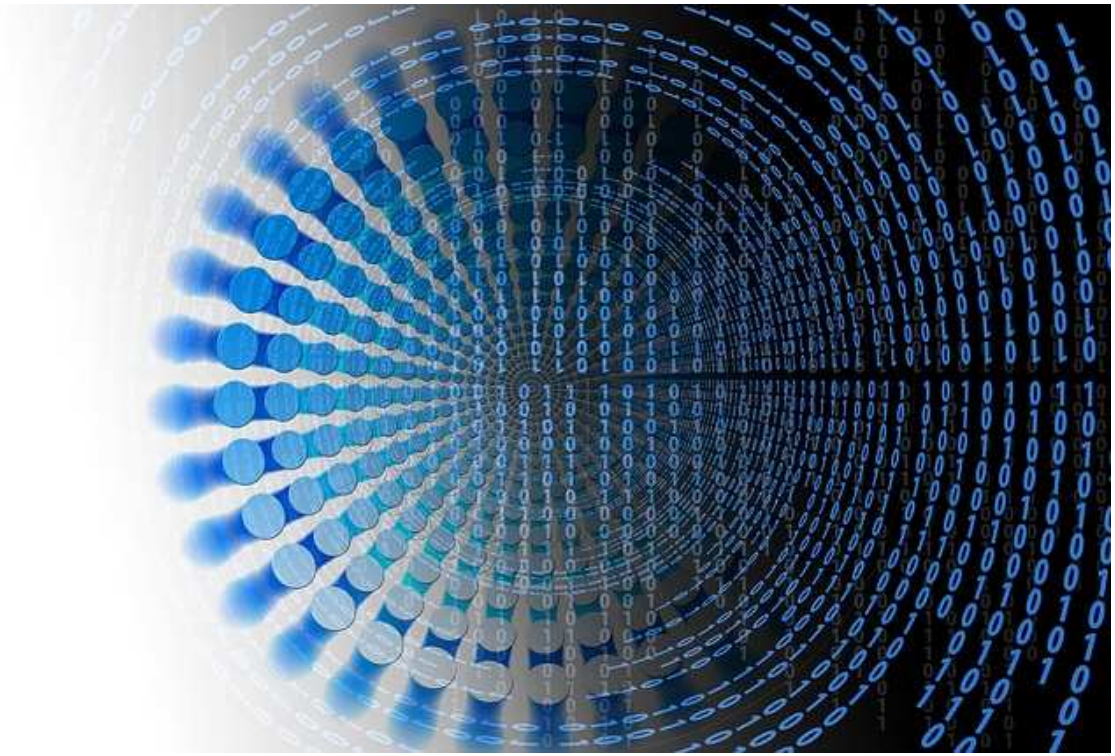
INDICE

PRINCIPALES DOCUMENTOS DE REFERENCIA





ANONIMIZACIÓN VS SEUDONIMIZACIÓN



GOVERTIS Advisory Services

AEC
ASOCIACIÓN ESPAÑOLA PARA LA CALIDAD

Seudonimización y anonimización no son sinónimos.

Seudoniminizar:

RGPD art 4.5)

«seudonimización»: el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado **sin utilizar información adicional**, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identificable;

Información adicional: 

Fuentes propias

Terceros

Datos públicos

Ejemplo seudonimización datos de calificaciones en un colegio

Nombre y apellidos	Materia	Calificación	Nombre y apellidos	Materia	Calificación
Alberto González	Inglés	4,5	1	Inglés	4,5
Eduard Chaveli	Inglés	6	2	Inglés	6
Samuel Parra	Inglés	5,5	3	Inglés	5,5

Anonimizar:

- **Diccionario de la lengua española de la Real Academia Española:**

Expresar un dato relativo a entidades o personas, eliminando la referencia a su identidad.

- **El RGPD no la define pero la menciona en el Considerando 26:**

Por lo tanto los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima, inclusive con fines estadísticos o de investigación.

- **Por lo tanto tenemos que entender que para que la información sea anónima NO debe tratarse de un dato de carácter personal:**

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

➢ **29WP Dictamen 05/2014 sobre técnicas de anonimización:**

A la luz de la Directiva 95/46/CE y de otros instrumentos jurídicos pertinentes de la UE, la anonimización es el resultado de un tratamiento de los datos personales realizado para evitar **de forma irreversible su identificación**.

Se ha demostrado que la anonimización irreversible no es posible

➢ **Orientaciones y garantías en los procesos de anonimización de la AEPD:**

La anonimización de datos debe considerarse como una forma **de eliminar las posibilidades de identificación de las personas**.

Ejemplo anonimización calificaciones colegio

Nombre y apellidos	Materia	Calificación		Materia	Nota media curso	% aprobados / suspensos
Alberto González	Inglés	4,5		Inglés	5,33	67% / 33 %
Eduard Chaveli	Inglés	6				
Samuel Parra	Inglés	5,5				

La seudonimización puede ser adecuada para:

- **Trasladar datos a un encargado de tratamiento**
- **Alojar o compartir datos en la nube**
- **Tratamiento de datos sensibles por varias áreas o departamentos de la misma empresa o AAPP**

La anonimización puede ser adecuada para:

- **Tratamientos de Big Data, Business Intelligence**
- **Tratamientos con fines históricos, estadísticos o científicos**
- **Tratamientos de inteligencia artificial**
- **Datos abiertos**

Cuestiones clave:

- ✓ La seudonimización comporta un tratamiento de datos personales y por ello le **aplica el RGPD y la LOPDGDD. Se trata de una anonimización limitada y reversible.**

- ✓ La anonimización de datos **NO comporta un tratamiento de datos personales** y por ello a los datos anonimizados **NO le aplica el RGPD y la LOPDGDD. Se pretende que sea IRREVERSIBLE.**

- ✓ El **proceso de anonimización constituye en sí un tratamiento de datos, por lo que a la hora de iniciar y realizar la anonimización aplica el RGPD.**

- ✓ Este proceso deberá por ejemplo basarse en alguna de **las bases jurídicas del tratamiento** y se **deberá informar al interesado** conforme al artículo 13 del RGPD si el tratamiento se realiza con fuentes endógenas o artículo 14 del RGPD si el tratamiento se realiza con fuentes exógenas.

- ✓**GT29:** para que exista una verdadera anonimización de datos personales, **ésta debe ser irreversible**, es decir, que razonablemente **NO permita la identificación del titular de los datos personales**. **Debido al estado de la tecnología lo que a priori no es identificable puede serlo aplicando nuevas tecnologías presentes o futuras.**

- ✓**GT29: Uno de los errores consiste en pensar que los datos seudonimizados son datos anonimizados. La probabilidad de identificación es muy alta por eso entra en juego la protección de datos ya que se sustituye un identificador único por otro.**

- ✓**Tras la anonimización** existe siempre un **riesgo de reidentificación de la información**, (lo que hoy es anónimo dentro de 5 años puede que no lo sea por el avance de la tecnología) lo que nos obliga a realizar **un análisis de riesgos** para eliminarlo o minimizarlo (riesgo residual aceptable), e incluso realizar una evaluación de impacto.

✓**Código de Buenas prácticas en protección de datos para proyectos de Big Data:**

- La **identificabilidad supone la aplicación de la normativa**
- Se refiere a que una persona pueda **ser identificada por un dato o por la combinación de información de diversas fuentes**.
- Para determinar si una persona **es identificable**, han de usarse **"todos los medios que puedan ser razonablemente utilizados y sin esfuerzos desproporcionados".**

✓ **Un buen ejemplo de seudonimización lo tenemos en la Ley de autonomía del paciente:**

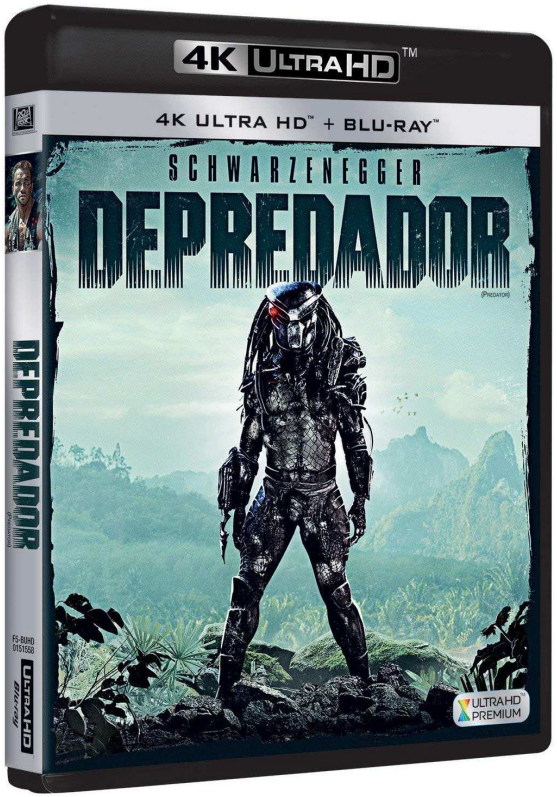
Art 16. 3. El acceso a la historia clínica con fines judiciales, epidemiológicos, de salud pública, de investigación o de docencia, se rige por lo dispuesto en la legislación vigente en materia de protección de datos personales, y en la Ley 14/1986, de 25 de abril, General de Sanidad, y demás normas de aplicación en cada caso. ***El acceso a la historia clínica con estos fines obliga a preservar los datos de identificación personal del paciente, separados de los de carácter clinicoasistencial, de manera que, como regla general, quede asegurado el anonimato, salvo que el propio paciente haya dado su consentimiento para no separarlos.***

✓ ¿Y si se reidentifican los datos anonimizados?

➢ Por cruce indebido con otras BBDD

➢ Por evolución de la tecnología

➢ Marketing quiere usar esos datos!!!!



✓**Que debe indicar el DPO:**

- **Supone la aplicación del RGPD y la LOPDGDD**
- **Se deben anonimizar los datos de nuevo**
- **En caso contrario y si la anonimización es intencional nos arriesgamos a sanciones tipificadas en la LOPDGDD**

Art 72.1. *En función de lo que establece el artículo 83.5 del Reglamento (UE) 2016/679 se consideran* ***muy graves*** *y prescribirán a los tres años las infracciones que supongan una vulneración sustancial de los artículos mencionados en aquel y, en particular, las siguientes:*

p) La reversión deliberada de un procedimiento de anonimización a fin de permitir la reidentificación de los afectados.

Otro tipo de infracciones en las que se podría incurrir

GRAVES	MUY GRAVES
La **falta de adopción de aquellas medidas técnicas y organizativas que resulten apropiadas para aplicar de forma efectiva los principios de protección de datos desde el diseño**, así como la no integración de las garantías necesarias en el tratamiento, en los términos exigidos por el artículo 25 del Reglamento (UE) 2016/679.	El tratamiento de datos personales sin que concurra alguna de las condiciones de **licitud del tratamiento** establecidas en el artículo 6 del Reglamento (UE) 2016/679.
La falta de adopción de las medidas técnicas y organizativas apropiadas para garantizar que, por defecto, solo se tratarán los datos personales necesarios para cada uno de los fines específicos del tratamiento, conforme a lo exigido por el artículo 25.2 del Reglamento (UE) 2016/679.	La utilización de los datos para una **finalidad que no sea compatible con la finalidad para la cual fueron recogidos**, sin contar con el consentimiento del afectado o con una base legal para ello.
El quebrantamiento, como consecuencia de la falta de la debida diligencia, de las medidas técnicas y organizativas que se hubiesen implantado conforme a lo exigido por el artículo 32.1 del Reglamento (UE) 2016/679.	**La omisión del deber de informar al afectado acerca del tratamiento de sus datos personales** conforme a lo dispuesto en los artículos 13 y 14 del Reglamento (UE) 2016/679 y 12 de esta ley orgánica.

Tanto con datos seudoniminizados como con datos anonimizados debemos tener presente que puede aplicar el artículo 11 RGPD:

Tratamiento que no requiere identificación

1. Si los fines para los cuales un responsable trata datos personales no requieren o ya no requieren la identificación de un interesado por el responsable, este no estará obligado a mantener, obtener o tratar información adicional con vistas a identificar al interesado con la única finalidad de cumplir el presente Reglamento.

2. Cuando, en los casos a que se refiere el apartado 1 del presente artículo, el responsable sea capaz de demostrar que no está en condiciones de identificar al interesado, le informará en consecuencia, de ser posible. En tales casos no se aplicarán los artículos 15 a 20, excepto cuando el interesado, a efectos del ejercicio de sus derechos en virtud de dichos artículos, facilite información adicional que permita su identificación.

EL PROCESO DE ANONIMIZACIÓN

GOVERTIS Advisory Services

AEC
ASOCIACIÓN ESPAÑOLA PARA LA CALIDAD

Las orientaciones de la AEPD recomiendan definir una **política de anonimización** que se encuentre documentada y actualizada, que contemple, al menos, los siguientes parámetros o elementos:

- Identificación de activos implicados en el proceso de anonimización.
- Equipo de trabajo asignado y segregación de funciones, atendiendo a perfiles o roles en relación con el proceso de anonimización, en línea con el principio de independencia profesional.
- Realización de una EIPD.
- Revisión de riesgos en caso de cambios en los procesos de anonimización y reevaluaciones periódicas del riesgo residual existente con el objetivo de introducir parámetros de mejora de la calidad de los procesos de anonimización.
- Formación e información al personal implicado en los procesos de anonimización con respecto al cumplimiento de la normativa de protección de datos personales, especialmente en relación con las medidas de seguridad de índole técnica y organizativa, la existencia y aplicación de una política de anonimización, medidas de control del personal con acceso a la información anonimizada, obligaciones y deberes en caso de ruptura de la cadena de anonimización y las actuaciones que debe realizar para paliar el impacto resultante de la materialización de alguno de los riesgos de reidentificación.

- Eliminación o reducción de variables que permitan la identificación de las personas cuyos datos se traten en las iniciativas de Big Data.
- Auditoría del proceso de anonimización y del uso posterior de los datos.

A la hora de iniciar un procedimiento de anonimización **NO bastaría con eliminar de una base de datos los campos de nombre, apellidos, dirección y DNI para considerar la información anonimizada,** puesto que si mantenemos por ejemplo los registros de código postal y fecha de nacimiento **es posible reidentificar al individuo sin métodos excesivos ni desproporcionados**.

Nombre y apellidos	DNI/NIE/ Pasaporte	Dirección	Código Postal	Fecha de nacimiento	Lugar de nacimiento	Profesión / oficio
		Amapola 3	28007	1949		
			28006	Marzo de 1967	Madrid	
		Amapola	28008			Estudiante
Mujer			28046	1947		

El 29WP analiza diferentes técnicas de anonimización en base a 3 riesgos:

- Que los registros permitan **singularizarse** en uno o varios interesados
- Que los registros puedan **vincularse** a uno o varios interesados
- Que del análisis de los registros algunos o todos puedan **inferirse** a un usuario

Técnicas:

- **Aleatorización:** modificar la veracidad de los datos a fin de eliminar el estrecho vínculo existente entre los mismos y la persona.
- **Adición al ruido:** modificar los atributos del conjunto de datos para que sean menos exactos.
- **Permutación:** mezclar los valores de los atributos en una tabla.
- **Privacidad diferencial:** generar vistas anonimizadas de un conjunto de datos, al mismo tiempo conservar una copia de los datos originales.
- **Agregación y anonimato k:** impedir que un interesado sea singularizado cuando se le agrupa junto con, al menos, un número k de personas.
- **Seudonimizar:** sustituir atributos por otros

- ✓ **Ninguna de las técnicas de anonimización mencionadas es 100% segura por lo que deben combinarse entre ellas**.
- ✓ **Se debe realizar el correspondiente análisis de riesgos del proceso de anonimización para posteriormente** gestionar los riesgos **resultantes de reidentificación** con medidas técnicas, organizativas o de cualquier otra índole.
- ✓ **La AEPD** destaca la utilidad de los algoritmos de cifrado, **resaltando los algoritmos de "hash"** como fórmula para garantizar la confidencialidad del dato No obstante, un mecanismo de hash **no garantiza por sí solo la irreversibilidad del dato**, es preciso **combinarlo con otras medidas tales como:**
  - **la aplicación de algoritmos de cifrado,**
  - **la utilización de sellos de tiempo,**
  - **la aplicación de capas de anonimización,**
  - **técnicas de reducción o técnicas de perturbación**.
- ✓ **A mayor distorsión de la información, menor calidad y fiabilidad.**

Retomamos el ejemplo anterior para analizarlo desde la perspectiva de la K – Anonimización:

Atributos clave o identificadores: nombre y apellidos y DNI

Cuasi-identificadores: el resto

Atributos sensibles: no existen

Nombre y apellidos	DNI/NIE/ Pasaporte	Dirección	Código Postal	Fecha de nacimiento	Lugar de nacimiento	Profesión / oficio
		Amapola 3	28007	1949		
			28006	Marzo de 1967	Madrid	
		Amapola	28008			Estudiante
Mujer			28046	1947		

Tomando como base los cuasi- identificadores podemos:

O bien generalizar los registros que están en rango

O bien reducir (eliminar) los registros que se encuentran fuera de rango **(en rojo)**

Dirección	Código Postal		Fecha de nacimiento		Lugar de nacimiento	Profesión / oficio
Amapola 3	28007	28***	1949	+ 65 años	Albacete	Abogado
Jorge Juan 6	28006	28***	Marzo de 1967	50 – 60 años	Madrid	Economista
Concha Espina 22	28003	28***	Febrero de 1999	20 – 30 años	Sevilla	Estudiante
Diagonal 57	08008	08***	Junio de 1968	50 – 60 años	Córdoba	Banca
Castellana 15	28046	28***	1947	+ 65 años	Valencia	Teleoperador
Princesa 23	28001	28***	1965	50 – 60 años	Santander	Profesor

La AEPD en su nota técnica de K-Anonimización menciona algunas herramientas optimas para realizar el proceso de anonimización.

MUCHAS GRACIAS POR SU ATENCIÓN.

GOVERTIS Advisory Services

AEC ASOCIACIÓN ESPAÑOLA PARA LA CALIDAD