LUXELLENCY

Anonimización de datos

Manual práctico de conceptos, herramientas y configuración en el marco del RGPD

AUTORA

Helena Debain

EDICIÓN

1 de agosto de 2026. Revisa la edición de 20 de junio de 2026.

CONTACTO

helenadebain@luxellency.com

Cada afirmación de este manual está contrastada con documentación oficial. Donde una fuente no ha podido verificarse, se dice.

Para qué sirve este manual

Cuando trabajas con datos de personas, una sola pregunta decide casi todo lo demás: ¿esos datos siguen permitiendo saber quién es cada individuo, o ya no? De la respuesta depende si el Reglamento General de Protección de Datos (RGPD) se aplica, qué obligaciones tienes encima y cuánto riesgo asumes al compartir, publicar o usar esa información para entrenar un modelo de inteligencia artificial.

Este documento está pensado para que puedas pasar de la teoría a la práctica sin tropezar. La primera parte define, una a una y con ejemplos numéricos, las nociones que necesitas dominar antes de abrir cualquier programa: qué es un cuasi-identificador, qué significa exactamente el k-anonimato, qué añaden la l-diversidad y la t-closeness, y qué es una jerarquía de generalización. Sin esos conceptos, configurar una herramienta es ir a ciegas; con ellos, la configuración se vuelve mecánica. La segunda parte recorre el proceso en cinco pasos que recomiendan las autoridades y, antes de ellos, el marco de comprobación que el Comité Europeo de Protección de Datos sometió a consulta pública en julio de 2026. La tercera compara las herramientas disponibles. La cuarta es el corazón del manual: la configuración detallada, pantalla a pantalla, de cada programa. Y las dos últimas abordan el terreno de la IA y el Reglamento europeo de inteligencia artificial, y cómo elegir según tu caso.

Todo lo que se explica aquí está contrastado con documentación oficial de las herramientas y de las autoridades de protección de datos. Las instrucciones de configuración reflejan el funcionamiento real de cada programa, no una aproximación.

Esta edición es de 1 de agosto de 2026 y revisa la de junio. Al final del documento encontrarás una nota con lo que ha cambiado y por qué, para que quien tenga la versión anterior sepa exactamente qué releer.

Licencia de uso

Parte 1. Conceptos que necesitas dominar antes de abrir un programa

1.1. Qué es un dato personal y los cuatro tipos de atributos

El RGPD considera dato personal toda información sobre una persona física identificada o identificable. La clave está en esa segunda palabra, identificable: no hace falta que aparezca el nombre para que el dato sea personal; basta con que la persona pueda ser señalada combinando piezas de información.

Por eso, lo primero al mirar un conjunto de datos es clasificar sus columnas. En cualquier tabla conviven cuatro tipos de atributos, y cada uno se trata de forma distinta.

Tipo de atributo	Qué es	Ejemplos	Qué se hace con él
Identificador directo	Señala a la persona por sí solo	Nombre y apellidos, DNI, número de teléfono, correo electrónico, número de la Seguridad Social	Se elimina o se sustituye por un seudónimo
Cuasi-identificador (identificador indirecto)	No identifica aislado, pero sí combinado con otros o con fuentes externas	Código postal, fecha de nacimiento, sexo, profesión, nacionalidad	Se transforma (generalización, supresión, etc.)
Atributo sensible	El dato de valor que se quiere conservar y que, si se asocia a una persona, puede causarle perjuicio	Diagnóstico médico, ingresos, ideología, orientación sexual	Se conserva, pero se protege con modelos como l-diversidad o t-closeness
Atributo no sensible	Información sin riesgo de privacidad	Identificador interno de producto, categoría genérica	Se conserva sin cambios

La mayor parte del trabajo de anonimización se concentra en los cuasi-identificadores, porque son los que reidentifican sin que lo parezca. El ejemplo clásico, documentado en la literatura académica: con solo el código postal, la fecha de nacimiento y el sexo se puede reidentificar a una proporción altísima de la población. Ninguno de los tres es un identificador directo, pero juntos reducen el universo de candidatos hasta dejar a una sola persona.

1.2. Anonimizar no es lo mismo que seudonimizar, y la diferencia es jurídica

Esta distinción no es un matiz de vocabulario. Marca la frontera de aplicación del RGPD.

La **anonimización** convierte los datos de modo que el individuo deja de ser razonablemente identificable, teniendo en cuenta el coste, el tiempo y la tecnología disponibles para intentar reidentificarlo. El **Considerando 26 del RGPD** establece que la información que no guarda relación con una persona identificada o identificable es información anónima, y que el Reglamento no se aplica a ese tipo de información. Dicho directamente: los datos verdaderamente anónimos quedan fuera del ámbito del RGPD. Esa es la razón por la que la anonimización es tan valiosa.

Conviene no leerlo a la ligera. Que el resultado quede fuera del RGPD no significa que el proceso lo esté. El tratamiento que realizas para llegar a esos datos anónimos, es decir, partir de datos personales y transformarlos, sí es una actividad sujeta al Reglamento. Y la anonimización real es exigente: no basta con borrar el nombre. Si la combinación de los datos restantes permite volver a señalar a una persona con un esfuerzo razonable, no has anonimizado, solo lo parece.

La **seudonimización** es otra cosa. El **Artículo 4.5 del RGPD** la define como el tratamiento de datos de manera que ya no puedan atribuirse a una persona concreta sin recurrir a información adicional, una clave que se guarda por separado. Como ese proceso es reversible (con la clave se vuelve al dato original), la información seudonimizada **sigue siendo dato personal** para quien conserva esa clave, y permanece sujeta a las garantías del RGPD.

La consecuencia práctica: si seudonimizas, reduces el riesgo y mejoras tu cumplimiento, pero sigues dentro del Reglamento. Si anonimizas de verdad, sales de su ámbito. Por eso la seudonimización se usa como medida de seguridad dentro de un tratamiento que continúa, mientras que la anonimización se reserva para cuando quieres liberar los datos de la carga regulatoria: publicarlos, cederlos o reutilizarlos con fines estadísticos o de investigación.

Hasta aquí, el planteamiento clásico. Desde 2025 hay una precisión que conviene incorporar, porque cambia la pregunta que hay que hacerse.

El Tribunal de Justicia de la Unión Europea, en su sentencia de 4 de septiembre de 2025 (asunto C-413/23 P, Supervisor Europeo de Protección de Datos contra Junta Única de Resolución), estableció que el carácter personal de un dato seudonimizado no es una etiqueta que viaje pegada al dato, sino que depende de los medios de los que razonablemente dispone cada sujeto que lo trata. El caso es fácil de recordar: la Junta Única de Resolución transmitió a Deloitte los comentarios de accionistas y acreedores del Banco Popular, identificados solo con un código alfanumérico y sin entregarle la clave de correspondencia. Para la Junta, que conservaba esa clave, eran datos personales. Para Deloitte, sin medios realistas de deshacer el código, podían no serlo.

Conviene precisar el alcance procesal de la sentencia, porque induce a error con facilidad. El fallo anula la sentencia del Tribunal General de 26 de abril de 2023 y devuelve el asunto T-557/20 a ese mismo tribunal, donde quedó archivado por auto de 19 de diciembre de 2025. Anular no equivale

aquí a rechazar el criterio: el Tribunal de Justicia respalda el enfoque relativo de la identificabilidad, que es precisamente el que después recoge el Comité Europeo de Protección de Datos en sus directrices de 2026.

De ahí se siguen dos cosas que conviene tener claras antes de ceder nada a un tercero.

La primera es que la pregunta correcta no es «¿esto está anonimizado?», sino «¿está anonimizado para quién?». La misma tabla puede ser dato personal en tus manos y no serlo en las de quien la recibe, o al revés si el destinatario tiene acceso a fuentes que tú no tienes.

La segunda es menos intuitiva y va en dirección contraria. Que los datos sean anónimos para el destinatario no te libera a ti. Si para ti eran datos personales en el momento de la cesión, la cesión fue una cesión de datos personales, y las obligaciones que llevaba asociadas, empezando por informar al interesado, siguen siendo tuyas. El Comité Europeo de Protección de Datos lo recoge expresamente en sus Directrices 02/2026 sobre anonimización.

1.3. k-anonimato: la definición exacta, con ejemplo

El **k-anonimato** es el modelo más extendido para medir el riesgo de reidentificación, y es el parámetro que vas a configurar en casi todas las herramientas. Por eso conviene entenderlo con precisión y no de oídas.

Un conjunto de datos cumple **k-anonimato** si, atendiendo a sus cuasi-identificadores, cada registro es indistinguible de al menos otros **k menos 1** registros. Dicho de otro modo: para cualquier combinación de valores de los cuasi-identificadores, existen como mínimo *k* personas que la comparten. A cada grupo de registros que comparten exactamente la misma combinación de cuasi-identificadores se le llama **clase de equivalencia**. Cumplir k-anonimato significa, simplemente, que no hay ninguna clase de equivalencia con menos de *k* registros.

Veámoslo. Partimos de esta tabla, donde Edad, Código postal y Sexo son cuasi-identificadores, y el Diagnóstico es el atributo sensible:

Edad	Código postal	Sexo	Diagnóstico
28	45003	M	Gripe
29	45004	M	Asma
47	28012	F	Migraña
46	28015	F	Gripe

Tal cual está, cada fila es única en sus cuasi-identificadores: cualquiera de esas combinaciones señala a una sola persona. Es un conjunto **1-anónimo**, es decir, sin protección. Aplicamos

generalización: la edad pasa a rangos de diez años y el código postal se recorta a sus tres primeras cifras.

Edad	Código postal	Sexo	Diagnóstico
20-29	450**	M	Gripe
20-29	450**	M	Asma
40-49	280**	F	Migraña
40-49	280**	F	Gripe

Ahora hay dos clases de equivalencia, cada una con dos registros idénticos en sus cuasi-identificadores. El conjunto es **2-anónimo**: ningún registro puede distinguirse de al menos otro. Si quisiéramos 3-anonimato, necesitaríamos que cada combinación la compartieran al menos tres personas, lo que exigiría generalizar más o suprimir registros.

Esos umbrales no son invención del sector. La guía básica de anonimización que publica la AEPD, traducción de la guía de la autoridad de protección de datos de Singapur, recomienda fijar un valor de **k más alto, por ejemplo 5 o más**, cuando los datos se van a intercambiar con terceros, y admite uno **más bajo, por ejemplo 3**, para el intercambio interno o la retención a largo plazo. La propia guía sitúa en 3 o 5 el umbral habitual del sector. La regla intuitiva: cuanto mayor es la exposición de los datos, mayor debe ser k.

Conviene retener dos matices. El primero, que la propia guía advierte de que no existen reglas rígidas: si no puedes alcanzar el umbral, la salida no es publicar igualmente, sino reforzar las medidas de seguridad. El segundo, que el Comité Europeo de Protección de Datos no fija ningún umbral numérico en sus directrices de 2026: un *k* alto es un indicio favorable, no una prueba, y no sustituye a la evaluación documentada que se describe en la Parte 2.

1.4. l-diversidad: cuando el k-anonimato no basta

El k-anonimato tiene un punto ciego. Imagina una clase de equivalencia con cinco personas (cumple 5-anonimato), pero en la que las cinco tienen el mismo diagnóstico sensible. Aunque no puedas saber **cuál** de las cinco es tu objetivo, sí puedes deducir su diagnóstico, porque todas lo comparten. A esto se le llama **ataque de homogeneidad**, y es una revelación de atributo, no de identidad.

La **l-diversidad** cierra ese hueco. Exige que, dentro de cada clase de equivalencia, el atributo sensible tenga al menos **l valores bien representados**. Si configuras una 3-diversidad, cada grupo de registros indistinguibles deberá contener al menos tres diagnósticos distintos, de modo que conocer la clase de equivalencia no permita deducir el dato sensible.

1.5. t-closeness: el siguiente nivel de protección

La l-diversidad también tiene límites: garantiza variedad, pero no que esa variedad sea representativa. Si en el conjunto global una enfermedad aparece en el 1 % de los casos pero en una clase de equivalencia concreta aparece en el 50 %, un atacante con conocimiento de fondo aprende algo, aunque haya diversidad.

La **t-closeness** lo resuelve exigiendo que la distribución del atributo sensible dentro de cada clase de equivalencia se parezca a su distribución en el conjunto completo, con una distancia máxima **t** entre ambas. Cuanto menor es *t*, más se parecen y menos información filtra cada grupo. Es el modelo más estricto de esta familia y el que mejor protege frente a ataques con información externa.

1.6. Privacidad diferencial: un enfoque distinto

Los tres modelos anteriores son **sintácticos**: trabajan sobre la estructura de la tabla. La **privacidad diferencial** es **semántica** y parte de otra idea. En lugar de generalizar, introduce ruido matemático calibrado en los resultados, de modo que la presencia o ausencia de un individuo concreto apenas cambie la salida. Su parámetro central es **épsilon (ε)**, que funciona como un presupuesto de privacidad: cuanto menor es épsilon, más ruido y más privacidad, pero menos precisión en los datos. Es el modelo de referencia cuando se publican estadísticas agregadas o se entrenan modelos con garantías formales, y herramientas como ARX la implementan en su variante (ε, δ).

1.7. Otros modelos que verás en las herramientas

Dos modelos más aparecen con frecuencia en los menús de configuración. La **δ-presencia** (delta-presencia) protege frente a un riesgo sutil: que un atacante averigüe si los datos de una persona están o no en el conjunto, lo que ya de por sí puede ser información sensible (por ejemplo, estar en una base de pacientes de una clínica concreta). El **k-map** es similar al k-anonimato, pero mide la indistinguibilidad contra una población externa de referencia, no solo contra el propio conjunto.

Algunas herramientas, como ARX, también te piden elegir un **modelo de atacante** al calcular el riesgo. Los tres habituales son el del **fiscal** (*prosecutor*: el atacante sabe que su objetivo está en el conjunto y va a por él), el del **periodista** (*journalist*: no sabe si está, pero intentará reidentificar a cualquiera) y el del **vendedor** (*marketer*: busca reidentificar al mayor número posible de registros, no a uno concreto). Elegir bien el modelo de atacante ajusta el cálculo del riesgo a tu escenario real.

1.8. Jerarquías de generalización: la pieza que de verdad hay que configurar

Aquí está el concepto que más cuesta a quien empieza, y el que más se usa al configurar ARX o Amnesia. Una **jerarquía de generalización** es la escalera de abstracción que defines para un cuasi-

identificador: dice cómo se va a ir generalizando un valor concreto, nivel a nivel, hasta su máxima abstracción.

Para un código postal, una jerarquía podría ser:

```
Nivel 0: 45003   (valor original)
Nivel 1: 4500*   (se oculta el último dígito)
Nivel 2: 450**   (se ocultan los dos últimos)
Nivel 3: 45***   (provincia)
Nivel 4: *****    (suprimido por completo)
```

Para una ciudad, la escalera sería semántica: Toledo → Castilla-La Mancha → España → Europa. Para la edad, intervalos: 28 → 25-29 → 20-29 → 0-99.

La herramienta no inventa estas escaleras: las defines tú, o le pides que las genere automáticamente. Después, el algoritmo prueba combinaciones de niveles (generalizar la edad al nivel 2 y el código postal al nivel 1, por ejemplo) hasta encontrar la que cumple tu k-anonimato perdiendo la menor información posible. Por eso, configurar bien las jerarquías es el 80 % del trabajo: de ellas depende que el resultado sea a la vez seguro y útil.

1.9. Supresión y límite de supresión

Generalizar no siempre basta. A veces, unos pocos registros atípicos (alguien con una combinación rarísima de cuasi-identificadores) impiden alcanzar el k-anonimato deseado sin tener que generalizar en exceso todo el resto. La **supresión** resuelve esto eliminando esos pocos registros conflictivos. El **límite de supresión** es el porcentaje máximo de filas que permites borrar. Las herramientas lo usan como válvula: en lugar de degradar todo el conjunto para acomodar cinco rarezas, sacrifican esas cinco filas y conservan la calidad del resto.

1.10. Tabla resumen de los modelos de privacidad

Modelo	Parámetro	Frente a qué protege	Idea en una frase
k-anonimato	**k**	Reidentificación (revelación de identidad)	Cada registro se confunde con al menos otros k-1
l-diversidad	**l**	Revelación de atributo por homogeneidad	Cada clase tiene al menos l valores sensibles distintos
t-closeness	**t**	Revelación de atributo con info externa	La distribución sensible de cada clase se parece a la global
Privacidad diferencial	**ε (épsilon)**	Inferencia sobre individuos	El resultado apenas cambia con o sin una persona
δ-presencia	**δ**	Saber si alguien está en el conjunto	Oculta la pertenencia al conjunto
k-map	**k**	Reidentificación contra una población	Indistinguibilidad medida contra un censo externo

Parte 2. Cómo se hace y cómo se comprueba

2.1. Antes de empezar: qué tiene que cumplir el resultado

Los cinco pasos que vienen después son el procedimiento. Pero conviene saber de antemano qué se comprueba al final, porque eso condiciona cómo se recorre el camino.

El Comité Europeo de Protección de Datos publicó el 7 de julio de 2026 sus **Directrices 02/2026 sobre anonimización**, sometidas a consulta pública hasta el 30 de octubre de 2026. No son todavía texto definitivo, pero ya marcan el criterio con el que se va a leer cualquier proceso de anonimización. Reformulan tres pruebas que venían del Dictamen 05/2014 del Grupo del Artículo 29 y las actualizan a la vista de la jurisprudencia reciente y de las técnicas de reidentificación basadas en IA.

En ese mismo pleno del 7 de julio de 2026 el Comité adoptó las **Directrices 03/2026 sobre extracción masiva de datos web en el contexto de la IA generativa**, también sometidas a consulta hasta el 30 de octubre. Van en paralelo a las de anonimización y conviene leerlas juntas si tu caso es alimentar o entrenar modelos.

Un conjunto de datos está anonimizado cuando se puede responder que no a las tres preguntas:

Criterio	La pregunta que hay que poder responder que no	Dónde se trabaja en este manual
Ausencia de singularización	¿Hay alguna combinación de valores que señale a una sola persona?	k-anonimato, sección 1.3
Ausencia de vinculación	¿Se puede enlazar un registro con otro conjunto que se refiera a la misma persona?	k-map y modelos de atacante, sección 1.7
Ausencia de inferencia	¿Se pueden deducir conclusiones concretas y significativas sobre una persona identificada?	l-diversidad y t-closeness, secciones 1.4 y 1.5

Hay además una decisión previa que el EDPB pide tomar de forma explícita y dejar documentada: con qué enfoque se evalúa.

El **enfoque contextual** mide lo que cada entidad concreta puede realmente hacer con los datos: qué información adicional tiene, de qué recursos dispone, a qué tecnología accede. Es el estándar legal completo y el que refleja la realidad. Su riesgo es que, si subestimas la capacidad del otro, te equivocas por el lado peligroso.

El **enfoque simplificado** ignora esas diferencias y asume que, si el medio de reidentificar existe en teoría, alguien acabará usándolo. Es más conservador, y el EDPB lo presenta como medida voluntaria de prudencia, no como una alternativa al estándar legal.

Lo práctico es encadenarlos. Pasa primero el filtro simplificado para localizar los agujeros teóricos y aplica después el análisis contextual a los escenarios de amenaza que en tu caso son realistas.

Dos ideas más conviene retener. La primera es que el listón no es riesgo cero, sino **probabilidad insignificante** de identificación: la exigencia es realista, no imposible. La segunda es que esa probabilidad aumenta con el tiempo, porque la tecnología mejora y van apareciendo conjuntos de datos nuevos con los que cruzar el tuyo. Por eso reevaluar de forma periódica no es una buena práctica opcional, sino parte de lo que se te va a pedir que demuestres.

2.2. El proceso en cinco pasos

La AEPD y el Comité Europeo de Protección de Datos (EDPB) describen la anonimización como un flujo de cinco pasos. La guía básica que la AEPD tradujo y publicó usa exactamente esta secuencia: conozca sus datos, desidentifique, aplique técnicas, calcule el riesgo y gestione el riesgo. Seguir el orden no es burocracia: es lo que te permite documentar después que el proceso fue razonable, que es justo lo que se te pedirá si alguien lo audita.

Paso 1. Conozca y clasifique sus datos. Identifica qué columnas son identificadores directos, cuáles cuasi-identificadores, cuáles atributos sensibles y cuáles no sensibles, según la tabla de la sección 1.1. Sin esta clasificación, ningún programa sabe qué transformar.

Paso 2. Desidentifique. Elimina los identificadores directos. Si necesitas conservar la capacidad de vincular registros del mismo individuo en el futuro, no borres sin más: asigna seudónimos (tokens o hashes) y guarda la tabla de correspondencia en un entorno seguro, cifrado y separado. Esa tabla es la clave del Artículo 4.5; mientras exista, estás en seudonimización, no en anonimización.

Paso 3. Aplique técnicas de anonimización. Trabaja sobre los cuasi-identificadores con las técnicas de la tabla siguiente.

Técnica	Qué hace	Ejemplo	Cuándo conviene
Supresión	Elimina columnas innecesarias o filas atípicas	Borrar la fila de un caso único e irrepetible	Outliers que impiden el k-anonimato
Enmascaramiento	Sustituye parte de un valor por símbolos	45003 → 450**	Códigos alfanuméricos (CP, matrículas)
Generalización	Reduce el detalle a categorías más amplias	Edad 28 → rango 25-29	El método principal para cuasi-identificadores
Perturbación	Añade ruido controlado a valores numéricos	Redondear ingresos a la centena	Datos numéricos donde importa la estadística agregada
Intercambio (*swapping*)	Reorganiza valores entre registros	Permutar códigos postales entre filas	Mantener distribuciones globales rompiendo el vínculo individual

Paso 4. Calcule el riesgo. Mide el k-anonimato resultante y, si procede, aplica l-diversidad o t-closeness sobre los atributos sensibles. Aquí decides el umbral (k=3 interno, k=5 externo como referencia) y compruebas si tu conjunto lo cumple.

Paso 5. Gestione el riesgo en el tiempo. La anonimización no es un acto único. El riesgo cambia: aparecen nuevas fuentes públicas, mejora la tecnología, se publican conjuntos que cruzados con el tuyo abren brechas. Por eso se acompaña de controles vivos: técnicos (cifrado, control de accesos), de proceso (auditorías, formación) y legales (acuerdos de no reidentificación con terceros). Documentar todo esto es la prueba de que actuaste con diligencia.

Esa documentación tiene ahora un contenido concreto. El EDPB pide dejar constancia del proceso seguido, de las pruebas realizadas y de sus resultados, conservarla después de haber anonimizado, y revisarla si ocurre un incidente de seguridad que afecte a la información que dabas por confidencial. Y pide algo más, fácil de pasar por alto: no calificar de anónimo un conjunto en el que la identificación sigue siendo posible. Llamar anónimo a lo que no lo es, en un contrato o en un aviso de privacidad, es en sí mismo un incumplimiento.

Parte 3. Las herramientas, comparadas

El RGPD no certifica ningún programa, pero sí marca la vara de medir: el resultado debe ser irreversible, el proceso debe contemplar los riesgos del contexto, el dato debe conservar utilidad y todo debe ser trazable. Con ese criterio, lo que de verdad diferencia a unas herramientas de otras es si aparecen o no referenciadas en la documentación de organismos europeos. Esta tabla resume el panorama; la configuración detallada de cada una está en la Parte 4.

Conviene decirlo con precisión, porque el vocabulario habitual exagera: ningún organismo europeo avala ni certifica herramientas de anonimización. Lo que sí puede comprobarse es si una herramienta aparece mencionada en documentación institucional, que es cosa distinta y bastante más modesta. Eso es lo que recoge la última columna, y solo eso.

Herramienta	Tipo de dato	Dificultad	Modelos de privacidad	Formato entrada	Sistema	Referencias institucionales
PDPC (vía AEPD)	Tabular simple	Muy baja	k-anonimato	Excel	Windows	Distribuida por la AEPD
ARX	Tabular	Baja-media	k-anon, l-div, t-clos, δ-presencia, k-map, dif.	CSV, Excel, BBDD	Multiplataforma (Java)	ENISA y EMA, atribución no confirmada (ver nota)
Amnesia	Tabular, set-valued	Baja	k-anon, k^m-anon	Texto delimitado (CSV)	Multiplataforma (Java)	OpenAIRE (financiación UE)
μ-Argus	Microdatos estadísticos	Media	Recodificación, supresión, ruido, microagregación	Microdatos + metadatos	Windows	Statistics Netherlands y Eurostat; citada por la AEPD junto a SUDA
sdcMicro	Microdatos estadísticos	Media-alta	k-anon, l-div, riesgo SUDA, microagregación	Cualquiera vía R	Multiplataforma (R)	Sin mención nominal localizada
Microsoft Presidio	Texto e imágenes	Alta (desarrollo)	Detección y transformación de PII por NLP	Texto, JSON, imágenes	Multiplataforma (Python)	Sin mención nominal localizada
FreeDPOTool	Tabular (CSV, Excel)	Muy baja	k-anonimato, hashing	CSV, Excel	Navegador Chrome	Sin mención nominal localizada

Un caso merece contarse en lugar de zanjarse, porque ilustra el problema entero. **ARX** aparece en esta tabla con una referencia que no he podido cerrar. La literatura académica sobre la herramienta, en concreto el artículo de Prasser y otros publicado en la revista Software: Practice and Experience en 2020, afirma dos cosas: que ARX figura en una guía de ENISA sobre métodos para aplicar la privacidad y la protección de datos desde el diseño, y que la Agencia Europea del Medicamento la ha mencionado como solución para la evaluación cuantitativa de riesgo en el marco de su Policy 0070.

No he conseguido localizar la cadena literal ARX en ninguno de esos dos documentos oficiales, ni en el informe de ENISA sobre privacidad desde el diseño ni en el de ingeniería de protección de datos de 2022. Dos revisiones independientes de este manual tampoco la encontraron.

Dejo las dos versiones a la vista en lugar de elegir una. La afirmación tiene respaldo en literatura revisada por pares y no tiene, a día de hoy y en mis manos, respaldo documental directo. Quien necesite esa referencia para justificar una elección de herramienta ante un auditor debería comprobarla por sí mismo antes de apoyarse en ella. Y si alguien localiza la página exacta, agradeceré el aviso: este documento se actualiza.

Parte 4. Configuración paso a paso de cada herramienta

4.A. Herramienta del PDPC distribuida por la AEPD (Excel)

Es la vía de entrada más rápida con respaldo de la autoridad española. La desarrolló la autoridad de protección de datos de Singapur (PDPC) y la AEPD la tradujo y publicó junto a su guía básica de anonimización, orientada expresamente a pymes y startups.

Descarga e instalación. Se descarga desde la web de la AEPD como archivo comprimido. El archivo está cifrado: al descomprimirlo con 7-Zip (o un programa equivalente) te pedirá una contraseña, que es **DIT-AEPD**. Dentro encontrarás la hoja de cálculo de Excel. No requiere instalar nada más: funciona sobre Excel en Windows. Si quieres comprobar la integridad del archivo descargado, desde la consola de Windows puedes calcular su huella con el comando certutil -hashfile indicando el algoritmo sha256. Un aviso honesto: en esta revisión no he localizado ninguna huella publicada por la AEPD contra la que contrastar el resultado, de modo que ese cálculo sirve para comparar dos descargas entre sí, no para validar el archivo frente a un valor oficial.

Uso, siguiendo los cinco pasos de la guía. 1. **Conozca sus datos.** Pega tu conjunto en la hoja e indica, para cada columna, si es identificador directo, cuasi-identificador o atributo sensible. 2. **Desidentifique.** La herramienta elimina o enmascara los identificadores directos que hayas marcado. 3. **Aplique técnicas.** Sobre los cuasi-identificadores aplica generalización y enmascaramiento con las reglas predefinidas que ofrece. 4. **Calcule el riesgo.** Calcula el nivel de k-anonimato del resultado y te orienta sobre el umbral adecuado según vayas a usar los datos internamente (k=3) o a divulgarlos (k=5). 5. **Gestione el riesgo.** Exportas el conjunto transformado y conservas el registro de las decisiones tomadas.

Su límite es su sencillez: solo Excel en Windows, conjuntos pequeños y técnicas básicas. Su ventaja es el respaldo explícito de la AEPD, que la convierte en la opción más defendible para trámites donde la documentación regulatoria importa.

4.B. ARX Data Anonymization Tool

Es la herramienta de referencia cuando se necesita solidez regulatoria y varios modelos de privacidad en un mismo flujo. Es gratuita, de código abierto, está escrita en Java y funciona en Windows, macOS y Linux. Se descarga desde arx.deidentifier.org.

El trabajo en ARX se organiza en **tres perspectivas** que verás como pestañas: **configurar**, **explorar** y **analizar**.

Fase 1. Configurar.

1. *Importa los datos*. ARX admite CSV, Excel y bases de datos. La tabla se carga en la perspectiva de configuración.
2. *Asigna el tipo a cada atributo*. Seleccionas una columna y le das uno de los cuatro tipos, que ARX distingue por colores en la cabecera: **rojo** para identificador (se eliminará), **amarillo** para cuasi-identificador (se transformará), **morado** para atributo sensible (se conserva, pero puede protegerse con l-diversidad o t-closeness) y **verde** para no sensible (se conserva intacto). Desde la versión 3.7 hay botones para asignar un tipo a todas las columnas de golpe.
3. *Define el tipo de dato*. En la pestaña de metadatos indicas si cada columna es texto (*String*), entero (*Integer*), decimal (*Decimal*), fecha (*Date/time*) u ordinal. Esto es importante: los asistentes de jerarquías solo funcionan bien si el tipo de dato es correcto.
4. *Crea las jerarquías de generalización*. Para cada cuasi-identificador defines su escalera de abstracción (sección 1.8). ARX ofrece cuatro asistentes según el tipo de variable: **basado en enmascaramiento** (ideal para códigos alfanuméricos como el código postal), **basado en intervalos** (para números, por ejemplo agrupar edades de diez en diez), **basado en orden** (para variables ordinales y cadenas con un orden lógico) y **basado en fechas** (para definir granularidad creciente: día, mes, año). Las jerarquías se pueden exportar y reutilizar en otros conjuntos con columnas similares.
5. *Selecciona y parametriza los modelos de privacidad*. En el panel correspondiente añades los modelos con el botón más: k-anonimato (e introduces tu *k*), y si has marcado algún atributo sensible, también l-diversidad o t-closeness. Puedes combinar varios a la vez.
6. *Ajusta el límite de supresión*. Hay un deslizador para el porcentaje máximo de registros que se pueden eliminar. La propia documentación de ARX recomienda dejarlo en **100 %**, para que el algoritmo tenga libertad de suprimir lo necesario y encuentre la mejor solución; también recomienda dejar desactivadas las opciones "Approximate" y de monotonía salvo que tengas un motivo de rendimiento.

Fase 2. Explorar. ARX no te da una única solución, sino el espacio de todas las combinaciones de niveles de generalización posibles, representado como un retículo (*lattice*). Antes, en el diálogo de anonimización, eliges la **estrategia de búsqueda**: óptima (garantiza la mejor calidad, pero puede ser lenta con conjuntos grandes), o heurística por número de pasos o por tiempo (más rápidas, sin garantía de optimalidad). También eliges entre transformación **global** (la misma regla para todos los registros) o **local** (reglas distintas para distintos subconjuntos). ARX resalta la transformación óptima; tú puedes seleccionar nodos del retículo y comparar.

Fase 3. Analizar. Aquí compruebas si la solución te sirve. ARX muestra dos análisis enfrentando datos de entrada y de salida: el de **utilidad** (cuánta información se ha perdido, con estadísticas de calidad) y el de **riesgo** de reidentificación, donde puede aplicar los modelos de atacante de fiscal, periodista y vendedor (sección 1.7). Si el equilibrio entre privacidad y utilidad te convence, exportas el conjunto anonimizado; si no, vuelves a la configuración y ajustas jerarquías o parámetros. La anonimización es, por naturaleza, iterativa.

4.C. Amnesia (OpenAIRE)

Es la herramienta del proyecto europeo OpenAIRE, y la única de esta lista con soporte nativo de **k^m-anonimato**, el modelo necesario cuando un individuo tiene asociado un conjunto de valores (por ejemplo, una lista de productos comprados o varios diagnósticos). Está escrita en Java y JavaScript. Hay una versión en línea para pruebas, pero los datos reales y sensibles deben anonimizarse con la **aplicación local descargable**, porque así ningún dato sale de tu equipo. Trabaja con archivos de texto delimitado (tipo CSV).

Los cinco pasos del proceso. 1. *Carga el archivo y clasifica*. Importas el fichero e indicas qué columnas son identificadores directos y cuáles cuasi-identificadores, además del tipo de dato de cada campo. 2. *Crea una jerarquía por cada cuasi-identificador*. Es obligatorio: cada cuasi-identificador necesita su escalera de generalización. Puedes construirla a mano o pedir a Amnesia que la **genere automáticamente**, y guardarla para reutilizarla. 3. *Elige el algoritmo y sus parámetros*. Seleccionas k-anonimato o k^m-anonimato, fijas el valor de *k* y enlazas cada jerarquía con su atributo. 4. *Ejecuta*. Pulsas "Execute" y Amnesia calcula las soluciones. 5. *Explora el espacio de soluciones y aplica*. Amnesia muestra todas las combinaciones de niveles como nodos: los **azules** son soluciones seguras (cumplen tu garantía) y los **rojos**, inseguras. Haces doble clic en un nodo azul para aplicar esa solución. Si una solución casi segura falla solo por unos pocos registros (por ejemplo, 5 de cada 100.000), en lugar de generalizar más todo el conjunto puedes **suprimir** esos pocos registros; Amnesia te muestra el porcentaje que supone. Finalmente, guardas el conjunto anonimizado en local.

4.D. sdcMicro (paquete de R)

Es la opción para entornos de análisis estadístico donde la reproducibilidad es un requisito de auditoría: todo el proceso queda escrito en código y se puede volver a ejecutar de forma idéntica. Exige saber R. Dispone además de una interfaz gráfica, **sdcApp**, para quien prefiera no programar.

Instalación y flujo básico. Desde R, se instala con `install.packages("sdcMicro")`. El flujo gira en torno a un objeto de la clase `sdcMicroObj`, que se crea con `createSdcObj()` declarando los cuasi-identificadores. A partir de ahí, las funciones de anonimización toman esa información automáticamente. Este es un ejemplo completo y funcional con los datos de prueba del propio paquete:

```
install.packages("sdcMicro")
library(sdcMicro)

# Datos de ejemplo incluidos en el paquete
data(testdata2)

# 1. Declarar los cuasi-identificadores (key variables)
kv <- c("urbrur", "roof", "walls", "water", "electcon", "relat", "sex")

# 2. Crear el objeto sdcMicro (con variable de ponderación muestral)
sdc <- createSdcObj(testdata2, keyVars = kv, w = "sampling_weight")

# 3. Obtener k-anonimato por supresión local, con k = 3
#    kAnon() es el nombre intuitivo de localSuppression()
sdc <- kAnon(sdc, k = 3)

# 4. Revisar el riesgo y cuántos valores se han suprimido
print(sdc)
```

La función `kAnon()` (equivalente a `localSuppression()`) suprime el mínimo de valores necesario para alcanzar el *k* indicado. Se pueden además aplicar otras técnicas como `microaggregation()` para variables numéricas, y exigir distintos niveles de *k* para distintas combinaciones de variables. Tras cualquier cambio manual, conviene recalcular el riesgo. Para trabajar con ventana gráfica en lugar de código, se lanza `sdcApp()`.

4.E. μ-Argus (ARGUS)

Es software de código abierto desarrollado por la oficina central de estadística de los Países Bajos (Statistics Netherlands) con financiación de Eurostat, y diseñado específicamente para proteger **microdatos estadísticos**. Es la herramienta de referencia en el contexto de la estadística oficial europea. Presupone conocimiento del dominio: no es una herramienta generalista.

Flujo general. μ-Argus trabaja a partir del fichero de microdatos acompañado de un fichero de metadatos que describe cada variable. Una vez cargados, se indican las variables identificadoras y se fijan los umbrales de riesgo; el programa localiza las combinaciones poco frecuentes (las que generan registros únicos o casi únicos) y permite tratarlas con sus métodos: **recodificación global** (equivalente a generalizar una variable en todo el conjunto), **supresión local** (ocultar valores concretos que siguen siendo arriesgados tras la recodificación), aleatorización, adición de ruido, microagregación e incluso generación de datos sintéticos. Finalmente se genera el fichero "seguro" para su publicación. La guía de la AEPD lo cita, junto al método SUDA, como herramienta para evaluar el riesgo en conjuntos donde el k-anonimato simple se queda corto, como los datos longitudinales o transaccionales.

4.F. Microsoft Presidio

Es un *framework* de código abierto para detectar y anonimizar información personal (PII) en **texto, imágenes y datos estructurados**, el terreno donde las herramientas tabulares no llegan. Está pensado para desarrolladores: se integra en código Python, no se usa con clics. Conviene retener una advertencia del propio proyecto: al basarse en detección automática, no garantiza encontrar toda la información sensible, por lo que debe complementarse con otras salvaguardas.

Presidio tiene dos motores. El **Analyzer** localiza las entidades personales (nombres, teléfonos, correos, etc.) usando reconocimiento de entidades nombradas, expresiones regulares y reglas, y devuelve para cada una su tipo, posición y una puntuación de confianza. El **Anonymizer** toma esos resultados y los transforma con un **operador**, que puede ser *replace* (sustituir por una etiqueta), *redact* (eliminar), *hash*, *mask* (enmascarar caracteres) o *encrypt* (cifrar, reversible con el motor de desanonimización).

Instalación y ejemplo.

```
pip install presidio-analyzer presidio-anonymizer
python -m spacy download en_core_web_lg
from presidio_analyzer import AnalyzerEngine
from presidio_anonymizer import AnonymizerEngine
from presidio_anonymizer.entities import OperatorConfig

analyzer = AnalyzerEngine()
anonymizer = AnonymizerEngine()

texto = "Contacta con John Smith en john.smith@example.com."

# 1. Detectar las entidades personales
resultados = analyzer.analyze(text=texto, language="en")

# 2. Anonimizar sustituyendo cada hallazgo por una etiqueta
salida = anonymizer.anonymize(
    text=texto,
    analyzer_results=resultados,
    operators={"DEFAULT": OperatorConfig("replace", {"new_value":
"<ANONIMIZADO>"})}
)
print(salida.text)
```

Por defecto Presidio trabaja en inglés. Para procesar texto en español hay que cargar un modelo de lenguaje español (por ejemplo `es_core_news_lg`), configurar el motor de NLP en consecuencia y pasar `language="es"` en la llamada al analizador.

4.G. FreeDPOTool

Es una extensión de Chrome que aplica k-anonimato, *hashing* y generalización directamente en el navegador, con procesamiento 100 % local. Acepta CSV y Excel. Su dificultad es muy baja: se instala desde la tienda de extensiones de Chrome y no requiere configuración previa. El flujo es directo: cargas el archivo, marcas qué columnas son cuasi-identificadores, fijas el valor de *k*, ejecutas y descargas el resultado, que la herramienta comprueba contra el k-anonimato indicado.

Conviene ser claro con su estatus. A diferencia de las anteriores, FreeDPOTool **no cuenta con ninguna referencia localizada en documentación oficial de ENISA, de la AEPD ni de ninguna autoridad europea de protección de datos**. Su propuesta de valor es funcional (privacidad desde el diseño, procesamiento local, sin dependencias externas) y para una prueba rápida o un perfil no técnico puede ser útil. Pero quien necesite documentar el respaldo institucional de la herramienta empleada debería apoyarse en alguna de las que sí aparecen citadas por organismos europeos.

Parte 5. Anonimización, modelos de IA y Reglamento europeo de IA

El terreno donde todo esto se vuelve más resbaladizo es el de la inteligencia artificial. Cuando se entrena un modelo con datos personales, el principio que más peso adquiere es el de **minimización**: no tratar más datos de los estrictamente necesarios. Si la finalidad se puede alcanzar con datos agregados o anonimizados, no hay justificación para usar datos personales en bruto.

5.1. Qué obliga el Reglamento de IA y desde cuándo

El **Reglamento (UE) 2024/1689**, conocido como Reglamento de IA o AI Act, se aplica por fases. Ese calendario cambió este mes: el **Reglamento (UE) 2026/1744, de 8 de julio de 2026**, publicado en el Diario Oficial de la Unión Europea el 24 de julio y en vigor desde el 27 de julio, aplaza parte de las obligaciones. Se le conoce como Ómnibus digital sobre IA.

Conviene leerlo al revés de como lo cuentan los titulares. Lo que se aplaza es lo que afecta a los proveedores de sistemas de alto riesgo. Lo que se mantiene es justamente lo que afecta a cualquier empresa que use IA de cara a sus clientes.

Obligación	Fecha de aplicación	Situación
Transparencia del artículo 50.1: avisar de que se está interactuando con un sistema de IA	2 de agosto de 2026	Se mantiene
Marcado legible por máquina del contenido sintético, artículo 50.2	2 de agosto de 2026, con transición hasta el 2 de diciembre de 2026 para los sistemas ya comercializados antes de esa fecha	Transición nueva
Nuevas prohibiciones: material íntimo no consentido y material de abuso sexual infantil generados con IA	2 de diciembre de 2026	Nuevas
Alto riesgo del anexo III, artículo 6.2: biometría, empleo, educación, migración, acceso a servicios esenciales	2 de diciembre de 2027	Aplazado
Alto riesgo del anexo I, artículo 6.1: sistemas integrados en productos con legislación armonizada de seguridad	2 de agosto de 2028	Aplazado
Espacios controlados de pruebas nacionales	2 de agosto de 2027	Aplazado

Traducido a lo que tiene que hacer un autónomo o una pyme, y conviene ser preciso porque circula una versión exagerada de esta obligación.

Si tienes un **chatbot** atendiendo a clientes, desde el 2 de agosto de 2026 tienes que avisar de que están hablando con una máquina, salvo que resulte evidente por el contexto. Si publicas **imágenes, audio o vídeo realistas** generados o manipulados con IA, contenido que pueda pasar por auténtico, tienes que identificarlo como sintético.

Con el **texto** la regla es más estrecha de lo que suele contarse. El artículo 50, apartado 4, solo obliga a declararlo cuando se publica con la finalidad de informar al público sobre asuntos de interés general, y aun en ese supuesto la obligación decae si el contenido ha pasado un proceso de revisión humana o control editorial y una persona física o jurídica asume la responsabilidad editorial de la publicación. Fuera de ahí no existe obligación general de etiquetar texto: ni un artículo de blog, ni una publicación en redes, ni unos materiales de curso, ni una campaña de marketing. Lo que no admite excepción es el tamaño de la empresa.

Y una precisión que evita un malentendido frecuente. La obligación de publicar un resumen del contenido usado para entrenar un modelo recae sobre quien pone ese modelo en el mercado, no sobre quien lo utiliza. Si tu empresa trabaja con un modelo de un tercero, esa obligación no es tuya. Las tuyas son la transparencia del artículo 50 y, si tratas datos personales, todas las del RGPD.

5.2. La alfabetización en IA: el texto legal y la nota de prensa no dicen lo mismo

Hay un punto del Ómnibus que conviene mirar con lupa, porque las dos fuentes oficiales no coinciden. Lo desarrollo entero, con fechas, porque es el ejemplo más limpio que conozco de por qué merece la pena ir al articulado en lugar de al resumen.

El punto de partida. El Reglamento (UE) 2024/1689 entró en vigor el 1 de agosto de 2024 y sus capítulos I y II se aplican desde el **2 de febrero de 2025**. Ahí vive el artículo 4. Dicho de otro modo: la obligación de alfabetización en IA lleva vigente desde hace año y medio, no llega ahora. Su redacción original exigía a proveedores y responsables del despliegue garantizar un nivel suficiente de alfabetización de su personal. El verbo era garantizar, y ahí estaba la fricción práctica: acreditar un resultado en cada persona es difícil de documentar y desproporcionado para una empresa pequeña.

Lo que propuso la Comisión. El **19 de noviembre de 2025** la Comisión Europea presentó el paquete Ómnibus digital. Su propuesta para el artículo 4 no lo reformulaba: suprimía la obligación directa sobre proveedores y responsables del despliegue y la sustituía por un mandato de fomento dirigido

a la Comisión y a los Estados miembros. La empresa dejaba de tener el deber; lo asumía la Administración.

Lo que ocurrió en la negociación. Según consta en la tramitación del procedimiento 2025/0359(COD), el Parlamento Europeo rechazó esa supresión en su posición de marzo de 2026 y propuso mantener la obligación, reformulada en términos de apoyo. Esa es la fórmula que recoge, casi literalmente, el texto final. El Consejo adoptó su mandato el 13 de marzo de 2026, el acuerdo se cerró en trílogo y el Reglamento (UE) 2026/1744 lleva fecha de 8 de julio de 2026. El Comité Europeo de Protección de Datos y el Supervisor Europeo emitieron su dictamen conjunto sobre este expediente el 20 de enero de 2026, dato que consta en el considerando 47 del propio Reglamento.

Lo que dice el texto aprobado. El nuevo artículo 4, en vigor desde el **27 de julio de 2026**, establece en su apartado 1 que proveedores y responsables del despliegue adoptarán medidas para apoyar la promoción de la alfabetización en materia de IA de su personal y de quienes se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso. Y añade una frase decisiva: esa obligación **no exige garantizar un nivel específico de alfabetización de ninguna persona en particular**.

Conviene detenerse en el sujeto de esa frase. El propio artículo llama obligación a lo que acaba de reformular. Y el considerando 8 del Reglamento explica que el artículo se modifica para exigir a proveedores y responsables del despliegue que adopten medidas de apoyo. Exigir, no sugerir.

Los apartados 2 y 3 reparten el resto del trabajo. La Comisión y los Estados miembros apoyarán y facilitarán los esfuerzos de las empresas, en particular de las pymes, y la Comisión publicará ejemplos prácticos de cumplimiento en la plataforma única de información. El Consejo de IA adoptará recomendaciones con objetivos comunes, teniendo en cuenta los marcos europeos de competencias; el considerando 8 cita expresamente DigComp, el Marco Europeo de Competencias Digitales para la ciudadanía.

Lo que dice la nota de prensa. El 27 de julio de 2026, el mismo día de la entrada en vigor, la Comisión Europea publicó en su portal de estrategia digital una nota que resume el cambio de otra manera: afirma que el anterior requisito de alfabetización en IA para las empresas queda sustituido por un estímulo no vinculante, con la Comisión y los Estados miembros asumiendo un papel más relevante en su promoción.

La contradicción, y por qué importa. Las dos frases no describen la misma norma. Una dice obligación reformulada; la otra dice estímulo no vinculante. Y la diferencia tiene consecuencias: si es obligación, tu empresa debe adoptar medidas y poder demostrarlo; si es estímulo, no debe nada. La explicación más razonable es que la nota resume la propuesta que la Comisión defendió en noviembre de 2025, no el texto que el Parlamento y el Consejo aprobaron en julio de 2026. Es un resumen desactualizado respecto de su propio expediente.

Qué hacer mientras tanto. Manda el articulado publicado en el Diario Oficial. Si tu empresa usa IA, sigues teniendo el deber de adoptar medidas de formación proporcionadas al conocimiento previo de cada persona, a su puesto y al contexto de uso del sistema, y sigues teniendo que poder acreditarlo. Lo que ha desaparecido es la exigencia de certificar un nivel individual, persona a persona. No es lo mismo que quedarse sin deber.

Y una observación del propio legislador que conviene retener, porque va más allá del cumplimiento. El considerando 8 afirma que la alfabetización en materia de IA debe ser una prioridad estratégica con independencia de las obligaciones reglamentarias y de las posibles sanciones. Dicho en corto: aunque no te obligaran, seguiría siendo lo sensato.

5.3. Datos sensibles para corregir sesgos: el nuevo artículo 4 bis

El Ómnibus introduce además un **artículo 4 bis** que interesa directamente a quien anonimiza, porque es el único punto donde la reforma amplía lo que puedes hacer con datos sensibles en lugar de restringirlo.

La norma habilita, con carácter excepcional, el tratamiento de categorías especiales de datos personales cuando sea **estrictamente necesario** para detectar y corregir sesgos. Los proveedores de sistemas de alto riesgo pueden ampararse en ella para cumplir las exigencias de calidad de los datos, y la posibilidad se extiende, en supuestos tasados, a responsables del despliegue y a sistemas y modelos que no son de alto riesgo, cuando el sesgo pueda afectar a la salud y la seguridad de las personas, incidir negativamente en derechos fundamentales o producir una discriminación prohibida por el Derecho de la Unión.

Las garantías son estrictas y conviene leerlas como condiciones, no como recomendaciones. Solo cabe acudir a esta base si el objetivo no puede alcanzarse con datos sintéticos o anonimizados. Los datos quedan sujetos a limitaciones técnicas de reutilización y a medidas de seguridad actualizadas. Y deben eliminarse una vez corregido el sesgo. El precepto aclara además que no crea ninguna obligación de detectar y corregir sesgos: abre una puerta, no impone recorrerla.

La lectura para este manual es directa. Antes de invocar el artículo 4 bis hay que haber comprobado que la anonimización o los datos sintéticos no sirven para el caso, y esa comprobación es exactamente el trabajo que describen las Partes 1 y 2.

5.4. Lo que no ha cambiado: el RGPD

Circula la idea de que la Unión Europea está desmontando el RGPD. A fecha de esta edición no ha ocurrido, y la confusión tiene una causa identificable: el paquete Ómnibus digital que la Comisión presentó el 19 de noviembre de 2025 contenía dos propuestas distintas, y solo una ha llegado a ser norma.

La primera es el **Ómnibus digital sobre IA**, procedimiento 2025/0359(COD), que es el que ha terminado en el Reglamento (UE) 2026/1744 descrito en las páginas anteriores. La segunda es el **Ómnibus digital general**, procedimiento 2025/0360(COD), que sí tocaría el Reglamento General de Protección de Datos, la Directiva de privacidad en las comunicaciones electrónicas, el Reglamento de Datos y la Directiva NIS2, e incluiría la definición de dato personal, el régimen de la seudonimización y las bases de legitimación para entrenar modelos.

Ese segundo expediente **no está aprobado**. A fecha de esta edición, el Observatorio Legislativo del Parlamento Europeo lo sitúa a la espera de decisión de comisión, es decir, en primera lectura y en fase de trabajo parlamentario. El Parlamento no ha adoptado posición en Pleno, no se han abierto negociaciones a tres bandas con el Consejo y no existe acuerdo. Constan el proyecto de informe de comisión de 22 de junio de 2026 y los dictámenes consultivos del Comité Económico y Social Europeo, del Banco Central Europeo y del Comité de las Regiones. Nada más.

Sobre el fondo, el Comité Europeo de Protección de Datos y el Supervisor Europeo de Protección de Datos adoptaron el 10 de febrero de 2026 un dictamen conjunto en el que expresan reservas de calado sobre la modificación de la definición de dato personal del artículo 4.1. Merece una precisión, porque de otro modo parece una contradicción: lo que discuten no es el enfoque relativo de la identificabilidad, que ellos mismos aplican en sus directrices de anonimización de julio de 2026, sino la conveniencia de fijarlo en el articulado del Reglamento con esa redacción concreta.

Mientras eso no se cierre, el marco aplicable es el que describe este manual. Si trabajas con datos personales, conviene seguir la tramitación, porque el resultado afectará directamente a la frontera entre seudonimización y anonimización que se explica en la sección 1.2.

5.5. El riesgo que se olvida: el dato puede salir del modelo

Hay un riesgo que el EDPB señaló expresamente en su **Dictamen 28/2024**, de 17 de diciembre de 2024, y que mucha gente pasa por alto: aunque un modelo no esté diseñado para devolver datos personales, puede acabar revelándolos ante determinadas preguntas. Un modelo entrenado sobre datos mal anonimizados puede, en ciertas condiciones, reproducir fragmentos de esa información. Por eso la evaluación de riesgos no debe quedarse en cómo se prepararon los datos antes del entrenamiento, sino contemplar también la posibilidad de que el dato personal se extraiga del modelo ya entrenado. Anonimizar bien la entrada es, en este contexto, también una forma de proteger la salida.

5.6. La AESIA: supervisar es solo una parte de su trabajo

Cuando se habla de la AESIA en España, casi siempre se la nombra como quien vendrá a sancionar. Es una lectura empobrecida y, sobre todo, incompleta.

La Agencia Española de Supervisión de la Inteligencia Artificial se creó por el **Real Decreto 729/2023**, que aprueba su estatuto. Junto a la supervisión y la potestad sancionadora, ese estatuto le

encomienda funciones de acompañamiento que rara vez se citan: promover entornos de prueba donde adaptar soluciones innovadoras al marco jurídico vigente, apoyar el desarrollo y el uso sostenible de la inteligencia artificial, impulsar un marco de certificación voluntario para entidades privadas, y crear conocimiento, formación y difusión. La propia Agencia se describe como el organismo responsable de la supervisión, el asesoramiento, la concienciación y la formación en materia de IA, dirigido tanto a entidades públicas como privadas.

Ese componente de apoyo no es teórico. España fue el primer país de la Unión Europea en poner en marcha un espacio controlado de pruebas de inteligencia artificial, establecido por el **Real Decreto 817/2023**. De las cuarenta y cuatro solicitudes presentadas se seleccionaron doce sistemas de alto riesgo, que recorrieron un itinerario de tres fases: formación y asesoramiento, análisis y adaptación, y acompañamiento al cumplimiento del Reglamento de IA. La continuación anunciada pasa por publicar guías técnicas y abrir nuevas convocatorias. La AESIA mantiene además un corpus de guías propias, entre ellas una guía introductoria al Reglamento de IA, y ha verificado la familia de modelos ALIA, la infraestructura pública española de modelos de lenguaje.

Para una pyme o un autónomo esto se traduce en algo concreto: antes de que exista una inspección hay una vía de consulta, unas guías publicadas y un espacio donde probar. Conviene usarla. Llegar a la relación con el supervisor el día en que pregunta por un incidente es la peor forma posible de empezarla.

Un apunte de reparto competencial. El Reglamento (UE) 2026/1744 refuerza las atribuciones de la Oficina Europea de IA sobre determinados sistemas construidos a partir de modelos de uso general y sobre los integrados en plataformas y motores de búsqueda de muy gran tamaño. Dónde termina esa competencia y dónde empieza la de las autoridades nacionales como la AESIA se lee en el articulado del propio Reglamento, y conviene consultarlo ahí y no en resúmenes de terceros.

Parte 6. Cómo elegir y buenas prácticas

La pregunta útil no es cuál es la mejor herramienta, sino cuál encaja con tu situación.

Tu situación	Herramienta recomendada
Uso operativo rápido, datos sencillos, trazabilidad ante la AEPD, sin perfil técnico	PDPC distribuida por la AEPD
Proyecto que exige solidez regulatoria y varios modelos de privacidad	ARX
Datos con atributos de tipo conjunto (listas asociadas a una persona)	Amnesia
Análisis estadístico en R con reproducibilidad auditable	sdcMicro
Estadística oficial o publicación de microdatos bajo estándares de Eurostat	μ-Argus
Texto libre, correos o imágenes, con automatización	Microsoft Presidio
Prueba rápida de un CSV sin instalar nada (sin respaldo institucional)	FreeDPOTool

Por encima de la elección concreta, tres hábitos sostienen cualquier proceso serio. Prioriza siempre la minimización: si la finalidad se alcanza con datos agregados, no trates datos personales. Mide el riesgo en lugar de confiar en la intuición, porque para eso existen el k-anonimato y sus refinamientos. Y documenta cada decisión, los parámetros usados y los accesos a las tablas de correspondencia, porque el día que alguien pregunte, esa documentación es la diferencia entre haber actuado con diligencia y solo afirmarlo.

Fuentes e instituciones de referencia

Reglamento General de Protección de Datos (RGPD), Considerando 26 y Artículo 4.5. Agencia Española de Protección de Datos (AEPD): guía básica de anonimización (traducción de la guía de la PDPC de Singapur) y herramienta asociada, y guía de orientaciones y procedimientos de anonimización. Comité Europeo de Protección de Datos (EDPB): Directrices 02/2026 sobre anonimización, Directrices 01/2025 sobre seudonimización y Dictamen 28/2024 sobre modelos de IA. Agencia Española de Supervisión de la Inteligencia Artificial (AESIA): estatuto, entorno controlado de pruebas y guías de orientación. ENISA (Agencia de Ciberseguridad de la Unión Europea): directrices sobre privacidad y protección de datos desde el diseño. Agencia Europea del Medicamento: documentación sobre evaluación cuantitativa de riesgo en su política de apertura de datos de ensayos clínicos. Eurostat y Statistics Netherlands: financiación y desarrollo de µ-Argus, y estándares de confidencialidad estadística. OpenAIRE: infraestructura europea de ciencia abierta y herramienta Amnesia. Documentación técnica oficial de ARX (arx.deidentifier.org), Amnesia (amnesia.openaire.eu), sdcMicro (CRAN y guía SDC Practice) y Microsoft Presidio (microsoft.github.io/presidio).

Marco legal y autoridades de protección de datos

Reglamento General de Protección de Datos Reglamento (UE) 2016/679, **Considerando 26 y Artículo 4.5**. Texto consolidado en EUR-Lex (español):

https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679

AEPD Guía básica de anonimización (traducción de la guía de la PDPC de Singapur):

https://www.aepd.es/documento/guia-basica-anonimizacion.pdf

AEPD Herramienta de anonimización del PDPC (descarga; el ZIP va cifrado con la contraseña DIT-AEPD):

https://www.aepd.es/descargas/herramienta-anonimizacion-pdpc

AEPD Orientaciones y garantías en los procedimientos de anonimización de datos personales:

https://www.aepd.es/guias/guia-orientaciones-procedimientos-anonimizacion.pdf

EDPB Guidelines 01/2025 on Pseudonymisation (adoptadas el 16 de enero de 2025 y sometidas a consulta pública hasta el 28 de febrero de 2025). A fecha de esta edición siguen en versión de consulta: no consta versión final adoptada. Página oficial:

https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en. PDF: https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf

EDPB Directrices 02/2026 sobre anonimización, versión 1.0, adoptadas el 7 de julio de 2026 y sometidas a consulta pública hasta el 30 de octubre de 2026. Página oficial:

https://www.edpb.europa.eu/public-consultations/guidelines-022026-on-anonymisation_en

PDF:

https://www.edpb.europa.eu/system/files/2026-07/edpb_guidelines_202602_anonymisation_v1_en_0.pdf

EDPB Dictamen 28/2024 sobre determinados aspectos de protección de datos en el contexto de los modelos de IA, de 17 de diciembre de 2024:

https://www.edpb.europa.eu/documents/opinion-of-the-board-art-64/opinion-282024-on-certain-data-protection-aspects-related-to_en

Tribunal de Justicia de la Unión Europea, sentencia de 4 de septiembre de 2025, asunto **C-413/23 P**, Supervisor Europeo de Protección de Datos contra Junta Única de Resolución. El fallo anula la sentencia del Tribunal General de 26 de abril de 2023 y devuelve el asunto T-557/20 al Tribunal General, que lo archivó por auto de 19 de diciembre de 2025 (asunto T-557/20 RENV). Fichas de los asuntos en CURIA:

https://curia.europa.eu/juris/liste.jsf?num=C-413/23 y https://curia.europa.eu/juris/liste.jsf?num=T-557/20

Reglamento (UE) 2024/1689 por el que se establecen normas armonizadas en materia de inteligencia artificial:

https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

Reglamento (UE) 2026/1744, de 8 de julio de 2026 (Ómnibus digital sobre IA), publicado en el DOUE el 24 de julio de 2026 y en vigor desde el 27 de julio de 2026:

https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32026R1744

Ficha en el BOE: https://www.boe.es/buscar/doc.php?id=DOUE-L-2026-81147

ENISA, Privacy and Data Protection by Design: from policy to engineering:

https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

Informe complementario **Data Protection Engineering** (2022):

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

Agencia Europea del Medicamento (EMA), External guidance on the implementation of Policy 0070, referencia EMA/90915/2016, versión 1.5 de 14 de mayo de 2025 (publicación de datos de

ensayos clínicos; el enfoque cuantitativo de riesgo sitúa el umbral de riesgo máximo de reidentificación en 0,09, equivalente a un valor de k igual o superior a 11):

https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/clinical-data-publication/support-industry/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data

Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), publicado en el BOE de 2 de septiembre de 2023, referencia BOE-A-2023-18911. Entre los fines de la Agencia figura expresamente el fomento de entornos reales de prueba de los sistemas de inteligencia artificial:

https://www.boe.es/buscar/doc.php?id=BOE-A-2023-18911

Real Decreto 817/2023, por el que se establece un entorno controlado de pruebas para el ensayo del cumplimiento del Reglamento de IA, en particular respecto de los sistemas de alto riesgo. Localizable en el BOE por su número.

AESIA, sede electrónica y corpus de guías propias, entre ellas la guía introductoria al Reglamento de IA, y verificación de la familia de modelos ALIA:

https://aesia.digital.gob.es

EDPB y EDPS, Dictamen conjunto 2/2026 sobre la propuesta de Reglamento relativa a la simplificación del marco legislativo digital (Ómnibus digital), adoptado el 10 de febrero de 2026, con reservas sobre la modificación de la definición de dato personal del artículo 4.1 del RGPD:

https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-22026-proposal_en

Observatorio Legislativo del Parlamento Europeo, procedimiento 2025/0360(COD), Ómnibus digital general. Es la fuente donde comprobar en cada momento el estado real de tramitación del expediente que afectaría al RGPD:

https://oeil.europarl.europa.eu/oeil/en/procedure-file?reference=2025/0360(COD)

EDPB, Directrices 03/2026 sobre extracción masiva de datos web en el contexto de la IA generativa, adoptadas el 7 de julio de 2026 en el mismo pleno que las Directrices 02/2026 y sometidas a consulta pública hasta el 30 de octubre de 2026. Disponibles en el apartado de consultas públicas de edpb.europa.eu.

Comisión Europea, nota de entrada en vigor del Ómnibus de IA, publicada el 27 de julio de 2026 en el portal Configurar el futuro digital de Europa. Es la fuente de la afirmación de que el requisito de alfabetización queda sustituido por un estímulo no vinculante, contrastada en la sección 5.2 con el articulado:

https://digital-strategy.ec.europa.eu/es/news/ai-omnibus-enters-force

Observatorio Legislativo del Parlamento Europeo, procedimiento 2025/0359(COD), Ómnibus digital sobre IA. Recoge la tramitación completa del expediente, incluida la posición del Parlamento sobre el artículo 4:

https://oeil.europarl.europa.eu/oeil/en/procedure-file?reference=2025/0359(COD)

Prasser, F. y otros, sobre la herramienta ARX, en Software: Practice and Experience, 2020. Es la fuente académica que atribuye a ENISA y a la Agencia Europea del Medicamento la mención de ARX, atribución que no he podido confirmar en los documentos originales y que se explica en la nota de la Parte 3.

Documentación oficial de las herramientas

ARX Data Anonymization Tool: https://arx.deidentifier.org/

Configuración detallada: https://arx.deidentifier.org/anonymization-tool/configuration/

Amnesia (OpenAIRE): https://amnesia.openaire.eu/

https://amnesia.openaire.eu/about-documentation.html infraestructura OpenAIRE: https://www.openaire.eu/

sdcMicro, paquete de R (CRAN): https://cran.r-project.org/web/packages/sdcMicro/index.html

Guía SDC Practice: https://sdcpractice.readthedocs.io/

µ-Argus, Statistics Netherlands (CBS), con financiación de Eurostat. Página de CBS: https://research.cbs.nl/casc/mu.htm

Repositorio mantenido (sdcTools): https://github.com/sdcTools/muargus

Microsoft Presidio: https://microsoft.github.io/presidio/

Repositorio: https://github.com/microsoft/presidio

Nota de esta edición (1 de agosto de 2026)

Esta versión revisa la de 20 de junio de 2026. Si tienes la anterior, esto es lo que ha cambiado y lo que no.

Sección 1.2. Se añade el criterio de la sentencia del TJUE de 4 de septiembre de 2025 (asunto C-413/23 P). El carácter personal de un dato seudonimizado depende de los medios de cada sujeto que lo trata, y la anonimización deja de ser una propiedad absoluta del conjunto. La edición de junio afirmaba sin matiz que el dato seudonimizado es siempre dato personal. Se precisa además el alcance procesal del fallo: anula la sentencia del Tribunal General de 26 de abril de 2023, devuelve el asunto y este queda archivado en diciembre de 2025. Anular no equivale a rechazar el criterio, y conviene decirlo porque se presta a confusión.

Parte 2. Sección nueva (2.1) con los tres criterios de las Directrices 02/2026 del EDPB, adoptadas el 7 de julio de 2026 y en consulta pública hasta el 30 de octubre de 2026, y con la elección entre enfoque contextual y simplificado. Los cinco pasos del proceso no cambian: pasan a ser la sección 2.2. Se amplía el paso 5 con las exigencias de documentación.

Sección 1.3. Los umbrales k=3 y k=5 dejan de atribuirse a la práctica profesional y pasan a citarse desde su fuente oficial: la guía básica de anonimización publicada por la AEPD. Se mantiene el matiz de que el EDPB no fija umbrales numéricos y se añade la advertencia de la propia guía sobre la ausencia de reglas rígidas.

Parte 5. Reescrita y ampliada. Incorpora el calendario del Reglamento de IA tras el Reglamento (UE) 2026/1744, en vigor desde el 27 de julio de 2026, con el detalle de qué se aplaza y qué no. Precisa el alcance real del artículo 50: aviso obligatorio en chatbots, identificación del contenido sintético realista en imagen, audio y vídeo, y declaración del texto solo cuando se publica para informar al público sobre asuntos de interés general, con la excepción de la revisión humana y la responsabilidad editorial. Aclara también que el resumen de datos de entrenamiento corresponde al proveedor del modelo y no a quien lo usa. Se añaden tres secciones nuevas: la 5.2, que reconstruye la cronología completa del artículo 4 desde el 2 de febrero de 2025 hasta el 27 de julio de 2026 y confronta la literalidad del articulado con la de la nota de prensa de la Comisión en materia de alfabetización en IA; la 5.3, sobre el artículo 4 bis y el tratamiento de categorías especiales de datos para detectar y corregir sesgos; y la 5.6, sobre la AESIA como organismo de apoyo y no solo de supervisión. La antigua 5.2 pasa a ser 5.4 y la antigua 5.3 pasa a ser 5.5.

Fuentes. Entradas nuevas: Directrices 02/2026 y 03/2026 del EDPB, Dictamen 28/2024, Dictamen conjunto 2/2026 del EDPB y el EDPS, sentencia C-413/23 P con su auto de archivo posterior, Reglamento (UE) 2024/1689, Reglamento (UE) 2026/1744, Real Decreto 729/2023, Real Decreto 817/2023, sede de la AESIA y ficha del procedimiento 2025/0360(COD) en el Observatorio Legislativo del Parlamento Europeo. Se han retirado tres enlaces a fuentes secundarias que se habían colado en el apartado y se ha actualizado la referencia de la EMA a su versión 1.5.

Parte 3. La columna de respaldo pasa a llamarse "Referencias institucionales", porque ningún organismo europeo avala ni certifica herramientas de anonimización. Se retira la atribución de ARX a la AEPD y se añade una nota bajo la tabla que expone las dos versiones sobre las referencias de ENISA y la EMA a esa herramienta: la que sostiene la literatura académica y la que no he podido confirmar en los documentos originales. Las fórmulas de ausencia de aval pasan a redactarse como ausencia de mención localizada.

Sección 4.A. Se corrige la explicación de la verificación por huella SHA-256: la contraseña DIT-AEPD es oficial, pero no he localizado ningún valor de huella publicado por la AEPD contra el que contrastar la descarga, de modo que el cálculo solo sirve para comparar dos descargas entre sí. Se rehace además el párrafo entero, que había quedado con un fragmento sobrante de la redacción anterior.

Licencia. Se añade al principio del documento un apartado propio que explica las condiciones de la licencia Creative Commons BY-NC-SA con la que se publica.

Lo que no ha cambiado. Las definiciones de los modelos de privacidad (secciones 1.3 a 1.7), las jerarquías de generalización (1.8), el límite de supresión (1.9), las instrucciones de configuración (Parte 4) y los criterios de elección (Parte 6). El Considerando 26 y el artículo 4.5 del RGPD siguen vigentes en su redacción original.