Unión Internacional de Telecomunicaciones

# UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

# X.1148

(09/2020)

SERIE X: REDES DE DATOS, COMUNICACIONES DE SISTEMAS ABIERTOS Y SEGURIDAD

Aplicaciones y servicios con seguridad (1) – Seguridad en la web

## Marco del proceso de desidentificación para proveedores de servicios de telecomunicaciones

Recomendación UIT-T X.1148

RECOMENDACIONES UIT-T DE LA SERIE X

**REDES DE DATOS, COMUNICACIONES DE SISTEMAS ABIERTOS Y SEGURIDAD**

| | |
|---|---|
| REDES PÚBLICAS DE DATOS | X.1–X.199 |
| INTERCONEXIÓN DE SISTEMAS ABIERTOS | X.200–X.299 |
| INTERFUNCIONAMIENTO ENTRE REDES | X.300–X.399 |
| SISTEMAS DE TRATAMIENTO DE MENSAJES | X.400–X.499 |
| DIRECTORIO | X.500–X.599 |
| GESTIÓN DE REDES DE INTERCONEXIÓN DE SISTEMAS ABIERTOS Y ASPECTOS DE SISTEMAS | X.600–X.699 |
| GESTIÓN DE INTERCONEXIÓN DE SISTEMAS ABIERTOS | X.700–X.799 |
| SEGURIDAD | X.800–X.849 |
| APLICACIONES DE INTERCONEXIÓN DE SISTEMAS ABIERTOS | X.850–X.899 |
| PROCESAMIENTO DISTRIBUIDO ABIERTO | X.900–X.999 |
| SEGURIDAD DE LA INFORMACIÓN Y DE LAS REDES | |
| Aspectos generales de la seguridad | X.1000–X.1029 |
| Seguridad de las redes | X.1030–X.1049 |
| Gestión de la seguridad | X.1050–X.1069 |
| Telebiometría | X.1080–X.1099 |
| APLICACIONES Y SERVICIOS CON SEGURIDAD (1) | |
| Seguridad en la multidifusión | X.1100–X.1109 |
| Seguridad en la red residencial | X.1110–X.1119 |
| Seguridad en las redes móviles | X.1120–X.1139 |
| **Seguridad en la web** | **X.1140–X.1149** |
| Protocolos de seguridad (1) | X.1150–X.1159 |
| Seguridad en las comunicaciones punto a punto | X.1160–X.1169 |
| Seguridad de la identidad en las redes | X.1170–X.1179 |
| Seguridad en la TVIP | X.1180–X.1199 |
| SEGURIDAD EN EL CIBERESPACIO | |
| Ciberseguridad | X.1200–X.1229 |
| Lucha contra el correo basura | X.1230–X.1249 |
| Gestión de identidades | X.1250–X.1279 |
| APLICACIONES Y SERVICIOS CON SEGURIDAD (2) | |
| Comunicaciones de emergencia | X.1300–X.1309 |
| Seguridad en las redes de sensores ubicuos | X.1310–X.1339 |
| Seguridad de las redes eléctricas inteligentes | X.1330–X.1339 |
| Recomendaciones relacionadas con la PKI | X.1340–X.1349 |
| Seguridad en la Internet de las cosas (IoT) | X.1360–X.1369 |
| Seguridad en los sistemas de transporte inteligente (ITS) | X.1370–X.1379 |
| Seguridad de tecnología de libro mayor distribuido | X.1400–X.1429 |
| Seguridad de tecnología de libro mayor distribuido | X.1430–X.1449 |
| Protocolos de seguridad (2) | X.1450–X.1459 |
| INTERCAMBIO DE INFORMACIÓN DE CIBERSEGURIDAD | |
| Aspectos generales de la ciberseguridad | X.1500–X.1519 |
| Intercambio de estados/vulnerabilidad | X.1520–X.1539 |
| Intercambio de eventos/incidentes/heurística | X.1540–X.1549 |
| Intercambio de políticas | X.1550–X.1559 |
| Petición de heurística e información | X.1560–X.1569 |
| Identificación y descubrimiento | X.1570–X.1579 |
| Intercambio asegurado | X.1580–X.1589 |
| SEGURIDAD DE LA COMPUTACIÓN EN NUBE | |
| Visión general de la seguridad de la computación en nube | X.1600–X.1601 |
| Diseño de la seguridad de la computación en nube | X.1602–X.1639 |
| Prácticas óptimas y directrices en materia de seguridad de la computación en nube | X.1640–X.1659 |
| Aplicación práctica de la seguridad de la computación en nube | X.1660–X.1679 |
| Otras cuestiones de seguridad de la computación en nube | X.1680–X.1699 |
| COMUNICACIÓN CUÁNTICA | |
| Terminologías | X.1700–X.1701 |
| Generador de números aleatorio cuántico | X.1702–X.1709 |
| Marco de seguridad QKDN | X.1710–X.1711 |
| Diseño de seguridad para QKDN | X.1712–X.1719 |
| Técnicas de seguridad para QKDN | X.1720–X.1729 |
| SEGURIDAD DE LOS DATOS | |
| Seguridad de los macrodatos | X.1750–X.1759 |
| SEGURIDAD DE 5G | X.1800–X.1819 |

*Para más información, véase la Lista de Recomendaciones del UIT-T.*

# Recomendación UIT-T X.1148

## Marco del proceso de desidentificación para proveedores de servicios de telecomunicaciones


**Resumen**

Las organizaciones de telecomunicaciones obtienen, gestionan, utilizan y comparten datos sobre las personas, incluida información de identificación personal, motivo por el que utilizan técnicas de desidentificación de datos para proteger los datos de los usuarios. En la Recomendación UIT-T X.1148 se describe el marco del proceso de desidentificación con sus fases operativas y se especifican, sobre la base del modelo del ciclo de vida de los datos y las funciones de los interesados, modelos de liberación de datos y etapas de datos dentro del proceso de desidentificación para proveedores de servicios de telecomunicaciones.



**Historia**

| Edición | Recomendación | Aprobación | Comisión de Estudio | ID único* |
|---|---|---|---|---|
| 1.0 | ITU-T X.1148 | 2020-09-03 | 17 | 11.1002/1000/14249 |

**Palabras clave**

Anonimia-k, cercanía-t, desidentificación, diversidad-l, modelos de liberación, proceso de desidentificación, protección de IIP, titular de datos.


* Para acceder a la Recomendación, sírvase digitar el URL http://handle.itu.int/ en el campo de dirección del navegador, seguido por el identificador único de la Recomendación. Por ejemplo, http://handle.itu.int/11.1002/1000/11830-en.

## PREFACIO

La Unión Internacional de Telecomunicaciones (UIT) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones y de las tecnologías de la información y la comunicación. El Sector de Normalización de las Telecomunicaciones de la UIT (UIT-T) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

## NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

## PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB en la dirección http://www.itu.int/ITU-T/ipr/.

## ÍNDICE

**Página**

## Introducción

La rápida evolución de las tecnologías y servicios de la información y la comunicación por Internet ha provocado un crecimiento exponencial de la cantidad de datos generados, transmitidos y almacenados. Los datos proceden de muchas fuentes, no sólo sensores, cámaras o dispositivos de red, sino también páginas web, sistemas de correo-e o redes sociales, entre otras muchas. Los conjuntos de datos son cada vez más grandes y complejos y llegan tan rápido que los métodos y herramientas tradicionales de procesamiento de datos ya no valen. Resulta muy problemático analizar eficientemente los datos con un retardo tolerable. Para resolver estos problemas se está desarrollando el paradigma conocido como analítica de macrodatos.

Las organizaciones de telecomunicaciones obtienen, gestionan, utilizan y comparten datos sobre las personas, incluida información de identificación personal, motivo por el que utilizan técnicas de desidentificación de datos para proteger los datos de los usuarios. Las relaciones entre las partes participantes en el flujo de datos para el intercambio de datos determinan si la desidentificación de los datos se ha de realizar antes de su obtención, tras la obtención, pero antes del almacenamiento, o sólo antes de su compartición con la siguiente parte en el intercambio. Del mismo modo, los proveedores de servicios de telecomunicaciones deben ofrecer la desidentificación de los datos como un servicio oportuno, eficiente y seguro para los clientes de los datos.

# Recomendación UIT-T X.1148

## Marco del proceso de desidentificación para proveedores de servicios de telecomunicaciones

### 1 Alcance

En esta Recomendación se presenta el proceso de desidentificación basado en el modelo del ciclo de vida de los datos, se especifica el marco del proceso de desidentificación con sus fases operativas y las funciones de los interesados. Se abordan además los modelos de liberación de datos y las etapas de datos dentro del proceso de desidentificación y se presentan varios métodos y ejemplos de desidentificación en los Anexos y Apéndices.

En esta Recomendación no se tratan temas relacionados con la reglamentación.

### 2 Referencias

Ninguna.

### 3 Términos y definiciones

#### 3.1 Términos definidos en otros documentos

En la presente Recomendación se utilizan los siguientes términos definidos en otros documentos:

**3.1.1 datos agregados** [b-ISO/CEI 20889]: Datos que representan a un grupo de titulares de datos, por ejemplo, las propiedades estadísticas de ese grupo.

**3.1.2 anonimización** [b-ISO/CEI 29100]: Proceso por el cual la información de identificación personal (IIP) se altera irreversiblemente de manera que un titular de la IIP ya no puede ser identificado directa o indirectamente, ni por el controlador de la IIP a título individual ni en colaboración con ningún otro interesado.

**3.1.3 atributo** [b-ISO/CEI 20889]: Característica inherente.

**3.1.4 conjunto de datos** [b-ISO/CEI 20889]: Colección de datos.

**3.1.5 desidentificación** [b-ISO 25237]: Término general utilizado para todo proceso de reducción de la asociación entre un conjunto de datos de identificación y el sujeto de los datos (véase la cláusula 3.2.4).

**3.1.6 proceso de desidentificación** [b-ISO/CEI 20889]: Proceso de eliminación de la asociación entre un conjunto de atributos de identificación y el titular de los datos.

**3.1.7 técnica de desidentificación** [b-ISO/CEI 20889]: Método de transformación de un conjunto de datos con el objetivo de reducir la medida en que la información puede asociarse a un titular de datos concreto.

**3.1.8 conjunto de datos desidentificado** [b-ISO/CEI 20889]: Conjunto de datos resultante de la aplicación de un proceso de desidentificación.

**3.1.9 información desidentificada** [b-NISTIR 8053]: Registro del que se ha eliminado u oscurecido suficiente IIP para que la información restante no identifique a un particular y que razonablemente no puede considerarse que vaya a utilizarse para identificar a una persona.

**3.1.10 privacidad diferencial** [b-ISO/CEI 20889]: Modelo formal de medición de la privacidad que garantiza que la distribución de probabilidad del resultado de un análisis estadístico difiere, como mucho, en un valor determinado, esté o no representado un titular de datos en el conjunto de datos analizado.

NOTA – Más concretamente, la privacidad diferencial ofrece:

a) una definición matemática de la privacidad que establece que, para considerar que el resultado de cualquier análisis estadístico preserva la privacidad, los resultados del análisis del conjunto de datos original no pueden distinguirse del que se obtendría si se añade o suprime un titular de datos del conjunto de datos; y

b) una medida de la privacidad que permite controlar la pérdida de privacidad acumulada y la fijación de un límite superior (o "presupuesto") para las pérdidas. La definición formal es la siguiente. Sea ε un número real positivo y M un algoritmo aleatorizado aplicado al conjunto de datos. Se considera que el algoritmo M es privado diferencialmente en ε si para todos los conjuntos de datos D1 y D2 que difieren en un único elemento (es decir, los datos de un titular de datos) y todos los subconjuntos S del alcance de M, mml_m1, donde la probabilidad está determinada por el grado de aleatorización del algoritmo.

**3.1.11 identificador** [b-ISO/CEI 20889]: Conjunto de atributos de un conjunto de datos que permite la identificación unívoca de un titular de datos en un contexto operativo específico.

NOTA – Véase en el Anexo B cómo esta definición se relaciona con las utilizadas en otras normas.

**3.1.12 atributo identificador** [b-ISO/CEI 20889]: Atributo de un conjunto de datos que contribuye a identificar unívocamente al titular de datos dentro de un contexto operativo específico.

**3.1.13 interesado en la privacidad** [b-ISO/CEI 29100]: Persona física o moral, autoridad pública, agencia u otro organismo que puede afectar, verse afectado o considerarse afectado por las decisiones o actividades relacionadas con el procesamiento de información de identificación personal (IIP).

**3.1.14 pseudonimización** [b-ISO/CEI 20889]: Técnica de desidentificación que sustituye un identificador (o identificadores) de un titular de datos con un pseudónimo para ocultar la identidad de ese titular de datos.

**3.1.15 semiidentificador** [b-ISO/CEI 20889]: Atributo del conjunto de datos que, considerado junto a otros atributos del conjunto de datos, identifica al titular de los datos.

**3.1.16 registro** [b-ISO/CEI 20889]: Conjunto de atributos relativos a un único titular de datos.

**3.1.17 reidentificación** [b-ISO/CEI 20889]: Proceso de asociación de datos de un conjunto de datos desidentificados con el titular de datos original.

NOTA – En esta definición se incluye el proceso que determina la presencia de un titular de datos concreto.

**3.1.18 destacar** [b-ISO/CEI 20889]: Aislar del conjunto de datos los registros pertenecientes a un titular de datos mediante la observación de las características que se sabe identifican unívocamente a ese titular de datos.

**3.1.19 tercero** [b-ISO/CEI 29100]: Interesado en la privacidad distinto del titular de la información de identificación personal (IIP), el controlador de la IIP y el procesador de la IIP, y las personas físicas autorizadas a procesar los datos bajo la autoridad directa del controlador de la IIP o el procesador de la IIP.

**3.1.20 tercero fiable** [b-ISO/CEI 18014-1:2008]: Autoridad de seguridad, o su agente, en el que confían otras entidades en lo relativo a las actividades de seguridad.

**3.1.21 anonimia-k** [b-ISO/CEI 20889]: Modelo formal de medición de la privacidad que garantiza que para cada identificador del conjunto de datos hay una clase de equivalencia correspondiente que contiene al menos K registros.

**3.1.22 diversidad-l** [b-ISO/CEI 20889]: Modelo formal de medición de la privacidad que garantiza que para un atributo seleccionado cada clase de equivalencia tiene al menos L valores bien representados.

NOTA – La diversidad-l es una propiedad del conjunto de datos que fija un límite inferior garantizado, L, para la diversidad de valores compartidos por una clase de equivalencia para un atributo seleccionado.

**3.1.23 cercanía-t** [b-ISO/CEI 20889]: Modelo formal de medición de la privacidad que garantiza que la distancia entre la distribución de un atributo seleccionado de una clase de equivalencia y la distribución de ese atributo en toda la tabla no rebasa el umbral T.

NOTA – Se dice que una tabla tiene cercanía-t con respecto a un atributo seleccionado si todas las clases de equivalencia que contienen ese atributo tienen cercanía-t.

## 3.2 Términos definidos en la presente Recomendación

En la presente Recomendación se definen los siguientes términos:

**3.2.1 controlador de datos**: Interesado (o interesado en la privacidad), distinto de las personas físicas que utilizan los datos con fines personales, que determina los fines y medios de procesamiento de datos.

**3.2.2 procesador de datos**: Interesado que procesa datos en nombre de un controlador de datos y conforme a las instrucciones recibidas de éste.

**3.2.3 encargado de la protección de datos**: Persona nombrada por el controlador de la información de identificación personal (IIP) para garantizar, de manera independiente, el cumplimiento de los requisitos legislativos o reglamentarios en materia de privacidad.

NOTA – "controlador de IIP" es sinónimo de "controlador de datos".

**3.2.4 sujeto de datos**: Entidad a que se refieren los datos.

NOTA – "sujeto de datos" es sinónimo de "titular de IIP" y de "titular de datos".

**3.2.5 procesar**: En relación con la información o los datos, implica la obtención, el registro o la retención de información o datos, o la realización de una o varias operaciones en los datos o la información, entre ellas:

– la organización, adaptación o alteración de la información o los datos;
– la extracción, consulta o utilización de la información o los datos;
– la revelación de la información o los datos mediante su transmisión, divulgación o puesta a disposición por otros medios; o
– la alineación, combinación, bloqueo, supresión o destrucción de la información o los datos.

# 4 Abreviaturas y acrónimos

En la presente Recomendación se utilizan las abreviaturas y acrónimos siguientes:

DP Privacidad diferencial (*differential privacy*)

DPO Encargado de la protección de datos (*data protection officer*)

IIP Información de identificación personal

TTP Tercero fiable (*trusted third party*)

# 5 Convenios

Ninguno.

## 6 Aspectos generales del proceso de desidentificación

El objetivo del proceso de desidentificación es proteger la confidencialidad de los sujetos de datos. Dado que esos datos pueden contener información de identificación personal (IIP), antes y después de analizar los datos para extraer información significativa, el analista de datos debe observar algunas consideraciones de seguridad.

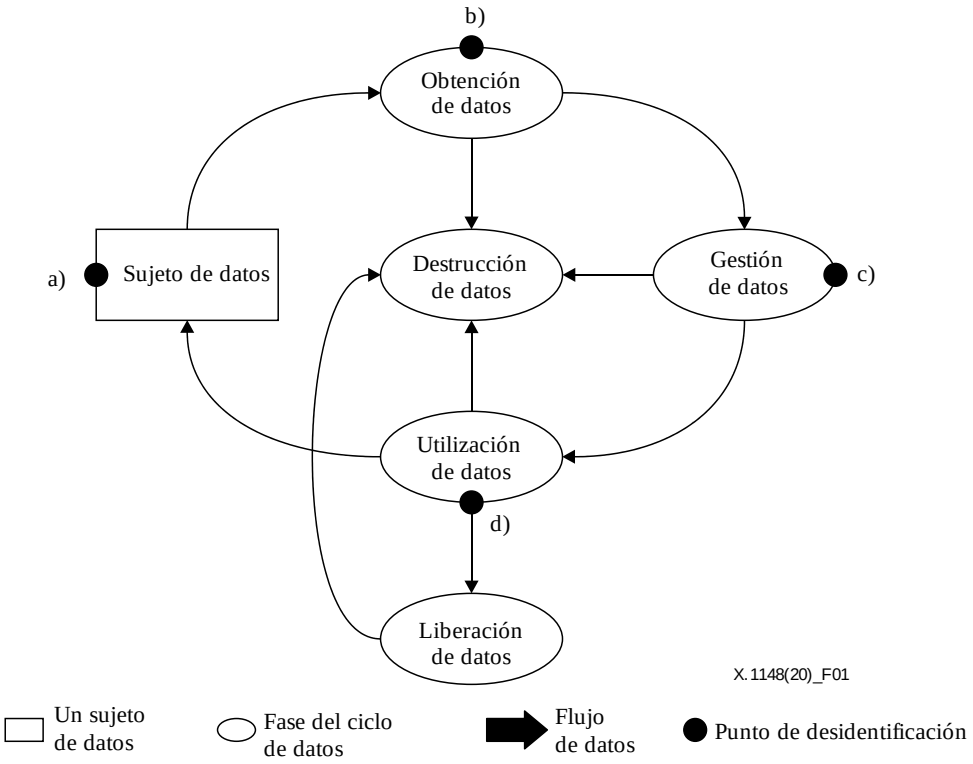
En esta cláusula se definen los entornos de análisis de datos, el modelo del ciclo de vida de los datos, las funciones de las entidades participantes en el proceso de desidentificación de datos y otras cuestiones relativas a la desidentificación.

### 6.1 Modelo del ciclo de vida de los datos y fase de desidentificación

Normalmente las organizaciones definen los objetivos de desidentificación con fines de privacidad y seguridad. En esta cláusula se define un ciclo de vida de los datos y se determina cuándo se ha de considerar un proceso de desidentificación basado en ese modelo de ciclo de vida de los datos.

El concepto de ciclo de vida de los datos se utiliza para seleccionar los controles adecuados en función de un análisis de las posibilidades de reidentificación. En esta Recomendación se define el ciclo de vida de los datos en las cláusulas 6.1.1 a 6.1.5.

En la Figura 1 se ilustra el proceso de desidentificación en el modelo de ciclo de vida de los datos.




* El sujeto de datos actúa como fuente de datos o receptor de datos.
La fase de gestión de datos comprende dos subfases de transformación y retención.

**Figura 1 – Proceso de desidentificación en el modelo de ciclo de vida de los datos**

#### 6.1.1 Fase de obtención de datos

Los datos se obtienen de los sujetos de datos, que son las personas a que se refieren los datos. El conjunto de datos producido como resultado de esta obtención puede contener IIP. La desidentificación crea un nuevo conjunto de datos del que se ha eliminado toda IIP. Se recomienda, siempre que sea posible, que las organizaciones utilicen a nivel interno conjuntos de datos desidentificados en lugar de los conjuntos de datos originales.

Con este modelo la desidentificación puede realizarse:

– durante la obtención de los datos, es decir, b) en la Figura 1; o

– cuando se han obtenido los datos sin real necesidad de un identificador, es decir, a) en la Figura 1.

No se deben obtener los identificadores que no sean necesarios para la gestión de los datos (transformación de datos y retención de datos).

### 6.1.2 Fase de gestión de datos

Para evitar el archivo de identificadores, la desidentificación debe aplicarse tras la transformación de los datos y antes de la retención de los datos, es decir, c) en la Figura 1. Se recomienda que las organizaciones consideren las posibilidades de reidentificación y establezcan controles de acceso, límites máximos de retención y políticas de supresión de datos claros para reducir al máximo los potenciales vínculos entre datos desidentificados. Se recomienda que las organizaciones consideren la posibilidad de aplicar técnicas de anonimización, como la agregación de datos, siempre que lo permita el objetivo para el que se utilizan los datos.

### 6.1.3 Fase de utilización de datos

Si una organización necesita IIP para gestionar los datos, se recomienda que estos se desidentifiquen antes de liberarlos como un conjunto para su compartición, es decir, d) en la Figura 1.

### 6.1.4 Fase de liberación de datos

Los datos se pueden compartir con terceros, limitados por controles administrativos adicionales, como acuerdos de "compartición de datos". Los conjuntos de datos desidentificados también pueden liberarse. La liberación de datos desidentificados se clasifica en tres categorías: pública, semipública o no pública. El nivel de desidentificación necesario puede variar en función del modelo de liberación escogido.

### 6.1.5 Fase de destrucción de datos

Los datos pueden destruirse en cualquier fase, a saber, la obtención de datos, la gestión de datos, la utilización de datos y la liberación de datos. Los datos deben destruirse con métodos verificados para evitar toda recuperación de los datos. Concretamente, se ha de considerar la necesidad de destruir los datos cuando se detecte que hay posibilidades de reidentificación.

## 6.2 Consideraciones acerca de la desidentificación

La aplicación de la desidentificación a lo largo del ciclo de vida de los datos aumenta su eficacia. Sin embargo, es la naturaleza de la relación entre las partes implicadas en el flujo de datos la que determina si la desidentificación de los datos debe realizarse antes de su obtención, es decir, a) en la Figura 1; tras su obtención, es decir, b) en la Figura 1, pero antes de su retención, es decir, c) en la Figura 1, o sólo antes de su compartición con la siguiente parte en el flujo de datos, es decir, d) en la Figura 1. Esta decisión influye a su vez en la viabilidad de las medidas de seguridad y de otro tipo de las organizaciones para mejorar la eficacia de la técnica de desidentificación utilizada en cada caso. Aunque la desidentificación puede resultar útil para proteger la confidencialidad de los datos de un sujeto cuando el objetivo para el que se utilizan los datos no soporta técnicas de anonimización, no es en sí misma suficiente para proteger los datos del sujeto y ha de considerarse como una parte del marco de protección de datos global. En esta cláusula se describen las características y consideraciones propias de cada fase.

### 6.2.1 Obtención de datos

El método más común es la desidentificación local (o desidentificación en origen), que permite a una persona (o controlador que procese los datos para una persona) eliminar toda la IIP antes de liberar los datos para su análisis.

Un aspecto de la desidentificación directamente relacionado con la fase de obtención de datos es la minimización de los datos. Un controlador de datos que obtenga datos de un sujeto deberá definir precisamente qué datos son estrictamente necesarios para el objetivo de utilización previsto y limitarse a obtener los datos que respondan exclusivamente a esos parámetros definidos.

A fin de reducir los campos de datos deberán implementarse procesos específicos para excluir la IIP innecesaria de la obtención/transferencia de datos.

Otro aspecto de la desidentificación es la agregación de datos. Los controladores de datos deben considerar la posibilidad de recurrir a la agregación de datos cuando su objetivo de uso no exige estrictamente la diferenciación de cada uno de los sujetos de datos.

### 6.2.2 Gestión de datos

#### 6.2.2.1 Transformación de datos

La fase de transformación de datos puede incluir la aplicación de técnicas de desidentificación como la agregación, la limitación de revelación estadística y la encriptación, entre otras. La transformación de los datos puede efectuarse en una o en varias etapas, incluso directamente tras su obtención y antes de la retención a largo plazo, tras un periodo de retención importante y antes del acceso, o integrarse en el acceso.

La transformación común de los datos mediante redacción o agregación puede efectuarse en cualquier momento entre la obtención y la liberación. Si se efectúa inmediatamente después de la obtención, la redacción o agregación de los datos puede reducir el potencial de daños para los sujetos de datos en caso de que haya una fuga de datos. Sin embargo, tras la redacción de los datos también se reducen las posibilidades de vinculación, fusión o actualización de los datos.

El método de transformación de los datos debe escogerse considerando detenidamente el daño que la divulgación de los datos podría causar a los sujetos de datos. Antes de tomar esa decisión también se han de tener en cuenta los análisis que deberán soportarse para utilizar los datos más adelante, pues las técnicas empleadas para reducir los riesgos de revelación pueden limitar las utilizaciones y análisis posteriores.

#### 6.2.2.2 Retención de datos

La retención de datos es el proceso de almacenamiento de datos, incluida la IIP, en cualquier forma de almacenamiento no volátil por el controlador de datos o una parte que actúe siguiendo sus instrucciones. Los controles de seguridad y privacidad de la información ya se centran en la fase de retención, por lo que en esta cláusula se resumen esos controles sin entrar en detalles [b-ISO/CEI 27001]. En la fase de retención se suele aplicar una serie de controles de seguridad y privacidad de la información, como el control de acceso, el mantenimiento, la evaluación de seguridad, los procedimientos de autentificación, la supervisión de incidentes y la correspondiente intervención, y las auditorías.

En particular, las organizaciones deben adoptar políticas de retención y eliminación de datos máximas para garantizar que los datos no se retienen más tiempo del estrictamente necesario para el objetivo de utilización correspondiente y que los datos se destruyen por completo una vez cumplido el periodo de retención máximo. Por ejemplo, los acuerdos de compartición de datos suelen especificar que el receptor debe destruir los datos al cabo de un tiempo especificado, por ejemplo, un año tras la recepción. Se trata de una condición contractual que puede estar impuesta por la legislación.

### 6.2.3 Utilización de datos

Los datos desidentificados pueden obtenerse, almacenarse o compartirse para diversos fines y aplicaciones, cada uno de ellos dependiente de ciertas propiedades de los datos conservadas tras la desidentificación. Uno de los principales motivos para liberar conjuntos de datos desidentificados es dar a otros la oportunidad de estudiar los valores y propiedades de los datos brutos con fines de investigación [b-ISO/CEI 20889]. Por consiguiente, otro de los objetivos de la desidentificación debe

ser la preservación de cuanta utilidad de la información sea posible, protegiendo al mismo tiempo la privacidad de las personas. Este doble objetivo de la desidentificación la convierte en un método importante cuya utilización puede considerarse en diversos contextos, incluidos los modelos de liberación de datos.

Al liberar datos desidentificados una organización debe tomar una decisión, generalmente a través de un comité de expertos formado por un amplio abanico de interesados, que considere las posibles consecuencias para los sujetos de los datos que se liberan. Para orientar esa evaluación y determinar el mecanismo de liberación adecuado que reduzca los riesgos de reidentificación se suele recurrir a la evaluación de riesgos y las listas de verificación.

La elección de las técnicas de desidentificación dependerá del grado de aplicabilidad o utilidad en cada caso concreto.

## 7 Marco del proceso de desidentificación

En esta cláusula se describe un marco del proceso de desidentificación para desidentificar la IIP en cuatro etapas, como se muestra en la Figura 2 [b-KOREA].

### Etapa 1 – Examen preliminar

La etapa 1 consiste en verificar si los datos objetivo son IIP o no. Si los datos no contienen IIP, se procede a la etapa 2. La desidentificación es necesaria.

### Etapa 2 – Desidentificación

La etapa 2 consiste en la desidentificación de los datos para evitar inferencias a partir de elementos concretos de información del conjunto de datos objetivo. En esta etapa se aplican métodos para eliminar o transformar total o parcialmente elementos de IIP. Entre los elementos de IIP se cuentan los identificadores, semiidentificadores y atributos sensibles.

### Etapa 3 – Evaluación de la adecuación

La etapa 3 consiste en evaluar la adecuación del conjunto de datos con elementos de IIP desidentificados. Se ha de considerar si el conjunto de datos objetivo sigue conteniendo IIP, si existe un potencial directo de reidentificación o si hay un potencial de vinculación que lleve a la reidentificación.

### Etapa 4 – Gestión del seguimiento

En la etapa 2 se mide la seguridad técnica y de gestión para evitar la reidentificación.

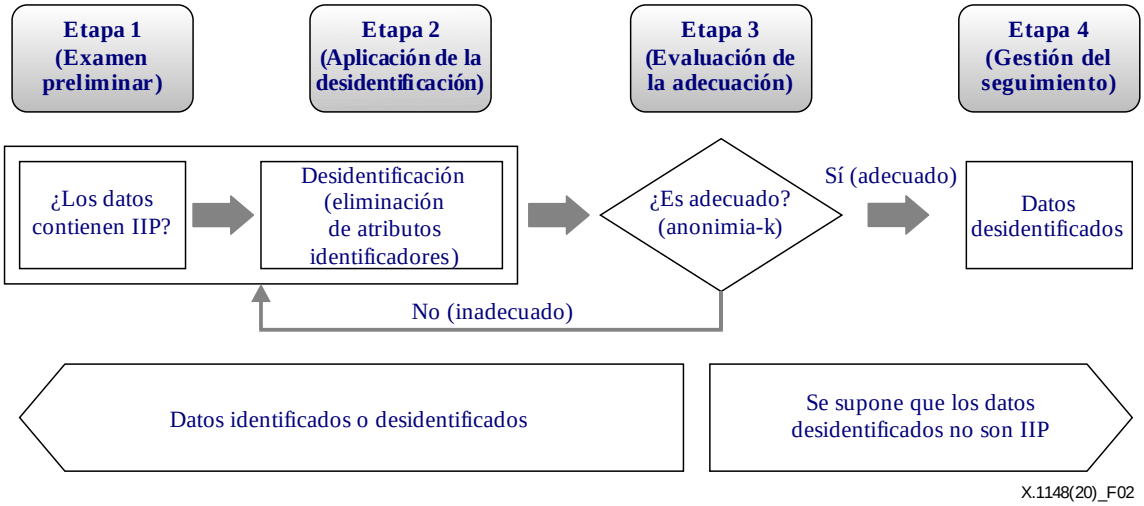



**Figura 2 – Proceso de desidentificación**

En las cláusulas 7.1 a 7.4 se detallan cada una de estas etapas.

### 7.1 Etapa 1 – Examen preliminar

Las organizaciones que pretendan utilizar o facilitar datos para diversos fines deberán, en primer lugar, determinar sus políticas y normas. Se recomienda que entre esas políticas y normas figuren las siguientes:

– cuáles son el objetivo y la utilización prevista de la información desidentificada;
– cuáles son los tipos de atributos de datos de los datos desidentificados;
– cuáles son las técnicas empleadas para la desidentificación;
– cuáles son los niveles de riesgo y efectos adversos de la reidentificación;
– cuáles son las soluciones disponibles en caso de reidentificación de una persona concreta;
– cómo se evalúa el nivel de reidentificación;
– cómo se determinan la mano de obra y el coste de la desidentificación.

Los elementos que hayan de considerarse específicamente en el examen preliminar pueden variar en función del tipo de datos y de su utilización prevista. No obstante, se recomienda fijar una serie de normas.

Las organizaciones que pretendan procesar datos para diversos fines se someterán a las normas correspondientes para verificar si los datos en cuestión son IIP o no. Incluso cuando se determine que los datos no son IIP, la organización deberá considerar si existe algún riesgo de vinculación entre los datos disponibles y tomar las medidas necesarias para minimizar ese riesgo. En caso de que los datos sean IIP, será necesario proceder a la desidentificación.

A continuación se ejemplifican los criterios de juicio de la IIP:

– no hay limitaciones especiales en los datos en términos de tipo, forma, características y formato;
– si el controlador de datos puede identificar a una persona con esos datos, éstos se considerarán IIP;
– los datos deben referirse a una persona. Un valor estadístico de un grupo formado por múltiples individuos no es IIP;
– los datos con los que se pueda identificar a una persona mediante su combinación con información adicional se consideran IIP. Por información adicional suele entenderse información pública o fácilmente disponible.

### 7.2 Etapa 2 – Aplicación de la desidentificación

#### 7.2.1 Desidentificación de identificadores

Un "identificador" es un dato, por ejemplo, un valor o un nombre unívocamente asignado a una persona o una cosa relacionada con una persona. Por norma general se ha de minimizar la obtención de "identificadores" y se han de suprimir todos los identificadores incluidos en los conjuntos de datos.

Sin embargo, los identificadores estrictamente necesarios para los fines previstos pueden ser datos como los siguientes:

– identificador unívoco (número de residencia, número de seguridad social, número de pasaporte, número de identificación de extranjeros, número de carnet de conducir, etc.);
– nombre (en caracteres chinos, nombre inglés, etc.);
– dirección completa (número de inmueble, calle, etc.);
– fechas (fecha de nacimiento, aniversarios (boda, etc.), fecha certificada, etc.);
– número de teléfono (móvil, fijo, profesional, fax, etc.);

– número de expediente médico, número de seguro sanitario nacional, número de beneficiario de prestaciones sociales, etc.;
– número de cuenta bancaria, número de tarjeta de crédito, etc.;
– fotografías (imágenes, vídeos, vídeos de televisión en circuito cerrado (CCTV), etc.);
– datos biométricos (huellas dactilares, voz, iris, etc.);
– dirección de correo-e, dirección IP, dirección de control de acceso a medios (MAC), localizador uniforme de recursos (URL) de la página principal, etc.;
– código de identificación (número de empleado, número de cliente, etc.);
– otros números de identificación unívoca (número de servicio militar, número de registro empresarial, etc.).

### 7.2.2 Desidentificación de semiidentificadores y atributos muy identificables

En general deben suprimirse los semiidentificadores de los conjuntos de datos si no valen para el fin a que se destinan los datos. Si un semiidentificador necesario para la utilización de los datos contiene elementos identificables deberán aplicarse técnicas de desidentificación como la pseudonimización y la agregación.

Los datos con un alto potencial de identificabilidad, como la información comportamental, deberán someterse a técnicas de desidentificación y, cuando sea posible, de anonimización.

### 7.2.3 Técnicas de desidentificación

Hay varias técnicas, como la pseudonimización, la agregación, la supresión de datos y la ocultación de datos, que pueden utilizarse individualmente o en combinación con otras. La pseudonimización puede no bastar por sí sola para la desidentificación.

Hay varias maneras de aplicar cada técnica. Deberá escogerse la más adecuada y utilizada en función del objetivo de utilización de los datos y de los puntos fuertes y débiles de cada técnica concreta. Una vez completada la desidentificación, se pasa a la etapa siguiente.

## 7.3 Etapa 3 – Evaluación de la adecuación para el proceso de desidentificación

Cuando la desidentificación no es suficiente se puede identificar a una persona combinando otros datos o utilizando técnicas de inferencia.

Para reducir el riesgo de reidentificación, se ha de proceder a una evaluación de la adecuación de los datos desidentificados. Esa evaluación debe considerar, entre otras cosas:

– cuál es el objetivo de la solicitud de desidentificación;
– qué tipo de atributos de datos participan en la desidentificación (si hay identificadores o no);
– cuál es el nivel adecuado de desidentificación.

La evaluación de la adecuación puede realizarla el encargado de la protección de los datos (DPO, *data protection officer*), un tercero fiable (TTP, *trusted third party*) delegado o un grupo de evaluación.

Al evaluar la adecuación se utiliza, entre otros modelos de protección de la privacidad, el modelo de anonimia-k. El modelo de anonimia-k es un método de valuación básico. De ser necesario, pueden utilizarse modelos de evaluación adicionales (diversidad-l, cercanía-t, privacidad diferencial (DP, *differential privacy*).

Véanse en el Anexo A más detalles sobre la evaluación de la adecuación.

## 7.4 Etapa 4 – Gestión del seguimiento

### 7.4.1 Medidas de protección para datos desidentificados

Se aplican medidas de protección para evitar la posibilidad de que se reidentifiquen los datos desidentificados en caso de que se combinen con otros datos y/o haya una fuga de datos. Esas medidas son, entre otras, las siguientes:

– medidas de protección administrativa: designación de la persona que se encargará de los ficheros de datos desidentificados, determinación de la compartición de los datos desidentificados y destrucción de los datos una vez cumplido su objetivo de utilización;

– medidas de protección técnica: restricción de acceso a los ficheros de datos desidentificados, gestión de los registros de acceso e instalación y ejecución de programas de seguridad.

Además, entre las medidas de seguridad también puede haber medidas de protección en caso de que haya una fuga de datos desidentificados. Entre esas medidas se cuentan las siguientes:

– análisis de la causa de la fuga y aplicación de medidas de seguridad administrativas y técnicas para evitar fugas ulteriores;

– retirada y destrucción de los datos desidentificados fugados.

### 7.4.2 Supervisión de las posibilidades de reidentificación

El controlador de datos que pretenda utilizar los datos desidentificados o facilitarlos a un tercero deberá supervisar periódicamente las posibilidades de reidentificación.

Cuando se detecte una posibilidad de reidentificación, se deberá solicitar al controlador de datos al que se han facilitado los datos desidentificados que suspenda su procesamiento, los retire y los destruya.

### 7.4.3 Requisitos para contratos con terceros

En todo contrato con terceros a los que se faciliten o confíen datos desidentificados para su utilización deberá incluirse la gestión del riesgo de reidentificación. La gestión del riesgo de reidentificación comprende:

– la notificación a los sujetos de datos de la revelación de los datos a terceros;

– siempre que sea posible, la facilitación a terceros de datos anonimizados;

– la prohibición de reidentificación: se estipulará que los controladores de datos a los que se han facilitado los datos, o encargado su procesamiento, tengan prohibida la reidentificación de los datos mediante su combinación con otros datos;

– la restricción de facilitación o entrega ulteriores: cuando se faciliten datos desidentificados o se encargue su procesamiento, se estipularán en el contrato los permisos de facilitación y procesamiento ulteriores;

– la notificación de riesgos de reidentificación: se estipulará la obligación de cesar el procesamiento de los datos e informar al consignador y el consignatario del problema de reidentificación cuando se haya detectado la reidentificación de los datos o las posibilidades de reidentificación sean elevadas.

### 7.4.4 Medidas contra la reidentificación

En caso de que se reidentifiquen los datos desidentificados, se deberá cesar su procesamiento y deberán tomarse las medidas necesarias para evitar la fuga de IIP.

Los datos reidentificados deberán destruirse inmediatamente.

# 8 Proceso de desidentificación de datos

## 8.1 Etapas de la desidentificación de datos

En esta cláusula se definen las etapas de desidentificación de datos, que pueden representarse como tipos de datos para describir el grado en que los datos identifican directamente a una persona y cómo esa persona está asociada a las características (atributos) de los datos. La especificación de los datos en el contexto de la utilización o el procesamiento de los datos debe incluir no sólo el tipo de datos, sino también una descripción del grado en que los datos pueden identificar a una persona o asociarla a una serie de características presentes en los datos.

En la Figura 3 se muestran las etapas, desde los datos identificados a los datos desidentificados, a lo largo del proceso de desidentificación. En cada etapa se corren diversos riesgos de reidentificación. Un tipo de datos especifica las etapas por las que pasará el conjunto de datos a medida que se va desidentificando progresivamente.

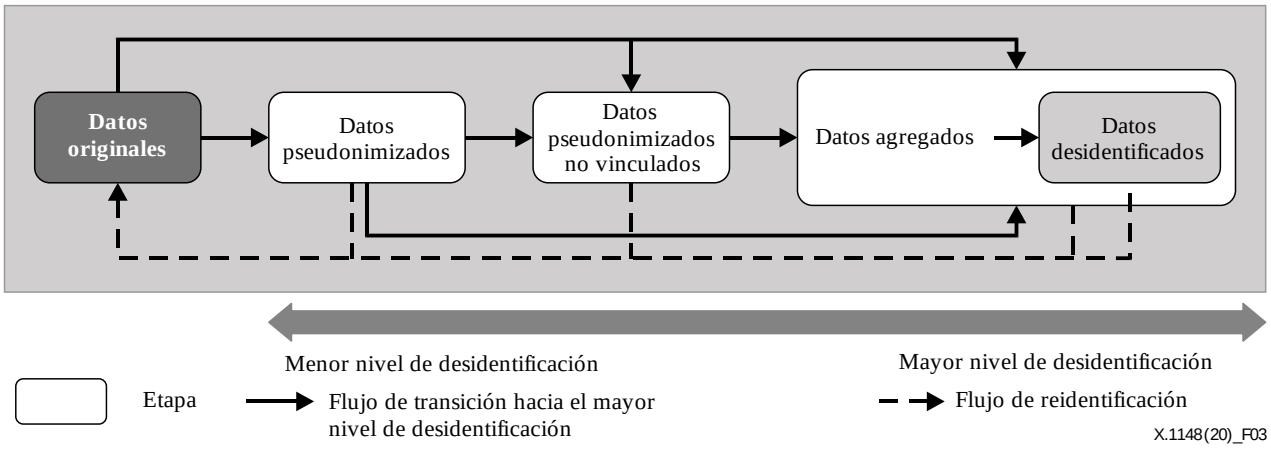



**Figura 3 – Etapas de desidentificación de datos**

Como se muestra en la Figura 3, en etapa de desidentificación están todos los datos. A la derecha (mayor nivel de desidentificación) se encuentran los datos desidentificados no relacionados con personas (por ejemplo, registros meteorológicos históricos), por lo que no hay riesgos para la privacidad. A la izquierda (menor nivel de desidentificación) están los datos identificados, vinculados directamente a personas concretas. Entre esos dos tipos de datos están los que pueden vincularse con esfuerzo, los que sólo pueden vincularse a grupos de personas y los que se refieren a individuos pero no pueden vincularse con ellos. En general, los procesos de desidentificación están diseñados para impulsar los datos hacia la derecha, conservando parte de la utilidad deseada, reduciendo el riesgo de distribuir datos desidentificados a una población más amplia o al público en general.

### 8.1.1 Datos originales

En la etapa de datos originales identificados, los datos pueden asociarse inequívocamente a una persona específica, pues su identidad figura en la información. Pueden consultarse las orientaciones sobre lo que puede considerarse como identificador en la cláusula 4.4.1 de [b-ISO/CEI 29100].

### 8.1.2 Datos pseudonimizados

En esta etapa ninguna persona distinta de la que ha asignado los alias puede, aplicando un esfuerzo razonable, revertir los datos, pues todos los identificadores se han sustituido por alias. Sin embargo, sigue siendo posible reidentificar los datos pesudonimizados vinculándolos con otros datos.

Estos datos corresponden a la definición de "pseudonimización" de la cláusula 3.1.14

### 8.1.3 Datos pseudonimizados no vinculados

En la etapa de datos pseudonimizados no vinculados se borran todos los identificadores o se sustituyen por alias, cuya función de asignación se borra o es irreversible, a fin de que nadie, incluida la parte que ha procedido a la operación, pueda con un esfuerzo razonable restablecer la vinculación. Sin embargo, sigue siendo posible reidentificar datos pseudonimizados no vinculados vinculándolos con otros datos.

### 8.1.4 Datos agregados

En esta fase los datos contienen información sobre un número suficiente de personas distintas de manera que no es posible inferir atributos individuales, pues los datos estadísticos no contienen entradas personales y están combinados. Con las técnicas de agregación, los datos agregados no alcanzan el nivel de identificabilidad inferior al umbral si la cantidad mínima de combinación de lagunas variables puede llevar a la identificación de una persona concreta.

Estos datos corresponden a la definición de "datos agregados" de la cláusula 3.1.1.

### 8.1.5 Datos desidentificados

En la etapa de datos desidentificados, los datos se desvinculan y los atributos se alteran (por ejemplo, se aleatorizan o generalizan los valores de los atributos) de manera que se puede confiar razonablemente en que los datos, por sí mismos o en combinación con otros datos, no basten para identificar directa o indirectamente a una persona.

## 8.2 Modelos de liberación de datos

El modelo de liberación de datos desidentificados se divide el tres modelos en función del contexto del análisis de los datos [b-UKAN].

Para la entrega de datos desidentificados se definen tres modelos de liberación: público, semipúblico y no público.

Cada modelo de liberación ofrece distintos niveles de disponibilidad y protección de la información. En función de los objetivos y/o los requisitos legislativos de la liberación de datos, variará la conveniencia de cada modelo. El modelo de liberación es una parte importante del proceso de desidentificación, pues la cantidad de desidentificación necesaria variará de acuerdo con el modelo de liberación seleccionado.

En las cláusulas 8.2.1 a 8.2.3 se exponen los tres modelos de liberación.

### 8.2.1 Modelo de liberación de datos público

Con la liberación de datos pública tradicional, cualquiera puede acceder a los datos sin necesidad de registrarse o cumplir condición alguna. Ejemplos de este tipo de liberación pueden ser los datos a disposición pública de las organizaciones y los datos publicados en depósitos de datos de acceso abierto, como los portales web. Las organizaciones liberan proactivamente conjuntos de datos y los ponen a gratuitamente a disposición de cualquiera para su utilización y reproducción.

Cuando los datos se liberan públicamente suelen imponerse el menor número de restricciones posible a la información, incluso en relación con quién y cómo puede acceder a ella. Así, cuando no es posible identificar a las personas que descargan el conjunto de datos, esa revelación debe considerarse una liberación de datos pública.

Al contrario de lo que ocurre con las solicitudes de acceso a la información de la cláusula 8.2.2, cuando no se exige que la persona que solicita la información se comprometa a cumplir determinados términos o condiciones en relación con el procesamiento, la privacidad o la seguridad de la información, se considera que se trata de una liberación de datos pública.

### 8.2.2 Modelo de liberación de datos semipúblico

El modelo de compartición de datos semipúblico es más restrictivo que el modelo de liberación de datos público y se da cuando se imponen una solicitud formal y un proceso de aprobación para poder acceder a los datos. En este caso el receptor de los datos podrá tener que comprometerse a cumplir determinados términos de utilización o firmar un contrato "mediante un clic". Los contratos mediante un clic son términos de utilización en línea que pueden limitar lo que se puede hacer con los datos y el tratamiento que éstos reciben. Aun así, cualquiera puede descargar los datos.

La desidentificación también puede resultar útil para responder a las solicitudes de acceso a la información de los conjuntos de datos. Al recurrir a la desidentificación, las organizaciones pueden responder a las solicitudes protegiendo la privacidad al tiempo que la utilidad de la información. Las organizaciones pueden utilizar controles de acceso para imponer ciertos límites cuando los datos se comparten a través de un sistema de información, por ejemplo:

– solicitar que todos los usuarios se registren y den sus datos de contacto antes de acceder a los datos;
– emplear protocolos de autentificación para verificar la identidad de una persona;
– utilizar sistemas de acceso progresivo para otorgar distintos niveles de acceso a distintas partes sobre la base, por ejemplo, de las afiliaciones o credenciales de la persona.

Con esos sistemas de información puede ponerse a disposición de una comunidad de investigadores un sistema de búsqueda interactivo, y los datos brutos pueden facilitarse a un pequeño número de analistas aprobados tras un detallado proceso de selección.

Asimismo, puede haber acceso a los datos sin necesidad de compartición de datos cuando los analistas solicitan que el controlador de datos realice un análisis en su nombre. Así, puede no haber compartición de datos por la organización.

### 8.2.3 Modelo de liberación de datos no público

Los conjuntos de datos que contienen IIP pueden compartirse dentro de las organizaciones, y entre ellas, sólo si la reglamentación vigente en el país permite su revelación. Si la revelación no está permitida y las instituciones siguen deseando compartir conjuntos de datos, se deberá eliminar toda la IIP. La liberación de datos no pública ofrece la menor disponibilidad, pero una mayor protección, y exige menos desidentificación.

Al compartir información entre organizaciones, dado que el acceso al conjunto de datos está limitado a la organización, se pueden definir requisitos de privacidad y seguridad de la información, cuyo cumplimiento se impone mediante acuerdos de compartición de datos. Para que la liberación de datos se considere no pública debe haberse concluido un acuerdo de compartición de datos entre las partes. El acuerdo de compartición de datos es una parte importante de la estrategia de reducción de riesgos de la liberación, y contiene términos comunes como los siguientes:

– especificación de quién tiene permiso de acceso (control del receptor);
– requisitos de seguridad de los datos (control de infraestructura);
– restricciones de uso, en particular la prohibición de vinculación con otros ficheros y de reidentificación deliberada (control de otros datos y control de gobernanza);
– requisitos de destrucción de los datos una vez utilizados (control de gobernanza).

Los acuerdos de compartición de datos cumplen tres objetivos:

– se distinguen claramente las personas y organizaciones en las que confía el controlador de datos y en las que no;
– se define el marco en que se especifican las condiciones de acceso;
– se pueden especificar las sanciones o penalizaciones aplicables en caso de incumplimiento de las condiciones por las personas/organizaciones.

### 8.2.4 Comparación de los modelos de liberación de datos

En un entorno de flujo de datos, una manera de limitar las posibilidades de reidentificación es controlar la manera en que los datos pueden obtenerse y utilizarse. Esos controles pueden clasificarse en función de distintos modelos de liberación de datos, cada uno de ellos con sus ventajas y riesgos inherentes. Las organizaciones también pueden optar por aplicar un método de acceso progresivo, que combina varios de los modelos para contemplar diversas posibilidades de uso y amenazas a la privacidad. Además, los modelos de liberación deben considerar la posibilidad de se efectúen liberaciones múltiples o periódicas. Los modelos definidos van desde los que no tienen restricción alguna a los que imponen restricciones firmes. En el Cuadro 1 se comparan los modelos de liberación de datos.

**Cuadro 1 – Comparación de los modelos de liberación de datos**

| | **Modelo de liberación público** | **Modelo de liberación semipúblico** | **Modelo de liberación no público** |
|---|---|---|---|
| Derechos de acceso | • Cualquiera puede acceder libremente a los datos liberados | • Pueden acceder a los datos liberados (o parte de los mismos) algunas personas u organizaciones | • Sólo un grupo de personas u organizaciones tiene acceso a los datos liberados |
| Casos de uso | • Acceso irrestricto a los datos mediante un portal web, es decir, libremente disponible para todos | • Configuración de seguridad del sitio<br>• Entrega de acceso<br>• Acceso a distancia virtual<br>• Acceso por análisis del servidor | • Compartición dentro de las organizaciones y entre ellas |
| Derechos | • Derechos ilimitados de reutilización y redistribución de los datos | • Disponible para las personas u organizaciones autorizadas | • Están prohibidas la reutilización, la reproducción o la distribución de los datos |
| Ataque de reidentificación | • Ataque de demostración con fines publicitarios | • Ataques internos deliberados<br>• Reconocimiento involuntario de una persona del conjunto de datos por conocimiento previo<br>• Fuga de datos | |

## 8.3 Relación entre modelos de liberación de datos y etapas de datos

### 8.3.1 Modelo de liberación de datos no público

Cuando se comparten datos de una fuente con el modelo de liberación no público, es necesario desidentificar los datos. En circunstancias normales, en lugar de utilizar el modelo de liberación no público, se utilizarán la pseudonimización no vinculada y una desidentificación de mayor nivel. En este caso, pueden utilizarse herramientas de desidentificación, como la pseudonimización, la criptografía, la síntesis, la supresión, etc.

Sin embargo, si las partes implicadas han concluido un contrato especial, los datos pseudonimizados pueden utilizarse para analizar y almacenar los datos durante esta fase.

### 8.3.2 Modelo de liberación semipúblico

Cuando se comparten datos de una fuente con el modelo de liberación semipúblico, se debe aplicar un nivel de desidentificación superior al del modelo de liberación no público. Se efectúa un procesamiento estadístico para prohibir la reidentificación. A continuación, pueden liberarse los datos

agregados y con un mayor nivel de desidentificación con el modelo de liberación semipúblico. Más concretamente, pueden utilizarse herramientas de desidentificación como la estadística, la aleatorización, etc.

Como se ve en el Cuadro 1, puede permitirse un nivel de desidentificación relativamente más bajo que con el modelo de liberación público, pues sólo algunas personas u organizaciones podrán acceder a los datos.

### 8.3.3 Modelo de liberación de datos público

Cuando se comparten datos de una fuente con el modelo de liberación público, se necesita un nivel de desidentificación superior al del modelo de liberación semipúblico. Se sigue el procedimiento de desidentificación de los datos y, al terminar, los resultados pueden utilizarse con el modelo de liberación público, como se muestra en el Cuadro 1.

# Anexo A

## Procedimiento de evaluación de la adecuación

(Este anexo forma parte integrante de la presente Recomendación.)

En este anexo se presenta un modelo de procedimiento de evaluación de la adecuación [b-KOREA]. Véase la Figura A.1.

A continuación se exponen las fases del procedimiento de evaluación de la evacuación:

– Preparación de los documentos básicos. El controlador de datos preparará los documentos básicos necesarios para evaluar la adecuación, por ejemplo, la declaración de datos, el nivel de desidentificación y el nivel de gestión de las organizaciones usuarias. Por "organización usuaria" se entiende la organización que pretende utilizar los datos desidentificados tras la desidentificación.

– Organización del grupo de evaluación. El encargado de la privacidad puede formar un grupo de evaluación o pedir al DPO o el TTP que realicen la evaluación.

– Desidentificación a partir de los documentos básicos preparados por el gestor de IIP.

– Aplicación de desidentificación adicional. El controlador de datos procederá a la desidentificación adicional en función de lo decidido por el grupo de evaluación en caso de que éste considere que la desidentificación no es adecuada.

– Utilización de los datos. Si se considera que la desidentificación es adecuada, los datos podrán utilizarse o facilitarse para, por ejemplo, el análisis de macrodatos.

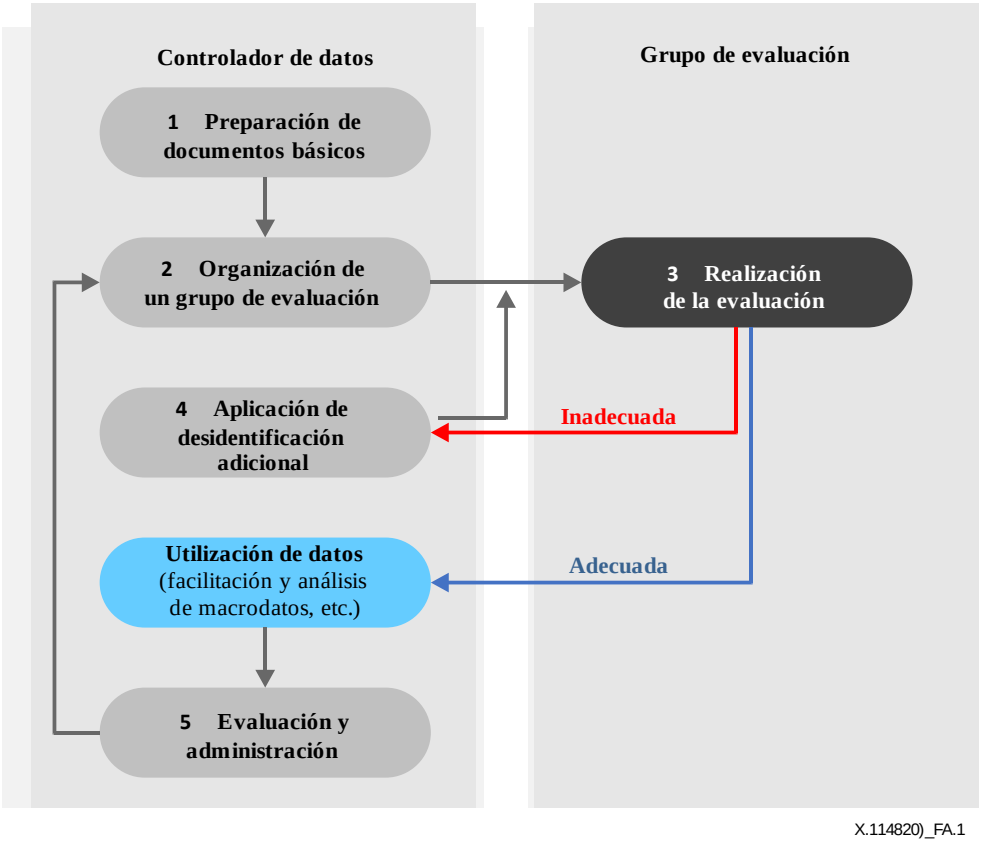



**Figura A.1 – Procedimiento de evaluación de la adecuación de la desidentificación**

## A.1 Preparación de los documentos básicos

El controlador de datos preparará los documentos básicos necesarios para evaluar la adecuación, como la declaración de datos de un sujeto de evaluación, el nivel de desidentificación y el nivel de gestión de la organización usuaria.

## A.2 Organización de un grupo de evaluación

El encargado de la privacidad puede formar un grupo de evaluación. Si se recurre a profesionales externos, se nombrará a más de un experto jurídico y en desidentificación a partir de la lista de expertos gestionada por las organizaciones especializadas en cada campo.

Los miembros del grupo de evaluación no estarán directamente interesados en el objetivo de utilización de los datos.

## A.3 Realización de la evaluación

El grupo de evaluación considera la adecuación de la desidentificación en función de los documentos básicos y utilizando el modelo de anonimia-k.

– Examen preliminar. Examen de los documentos básicos preparados por el controlador de datos y verificar en la entrevista si el conjunto de datos contiene elementos de identificación personal y si el objetivo de utilización y las técnicas de desidentificación son apropiados.
– Análisis de las posibilidades de reidentificación. Análisis de los posibles intentos de reidentificación, incluso voluntarios, del nivel de protección de la IIP y de la capacidad del controlador de datos que utiliza o recibe los datos.
– Análisis de consecuencias de la reidentificación. Evaluación de las posibles consecuencias para el sujeto de datos de la reidentificación voluntaria o involuntaria de los datos.
– Análisis cuantitativo. Verificación de la exactitud del valor K facilitado por el controlador de datos.
– Decisión del valor de referencia de la evaluación. El grupo de evaluación determina globalmente el valor de referencia de la evaluación habida cuenta de la reidentificación, las consecuencias de la reidentificación, los resultados del análisis cuantitativo y el objetivo de utilización de los datos.
– Evaluación de la adecuación. Se decide la adecuación de la desidentificación comparando los valores calculados resultantes del valor de referencia medio y del análisis cuantitativo.

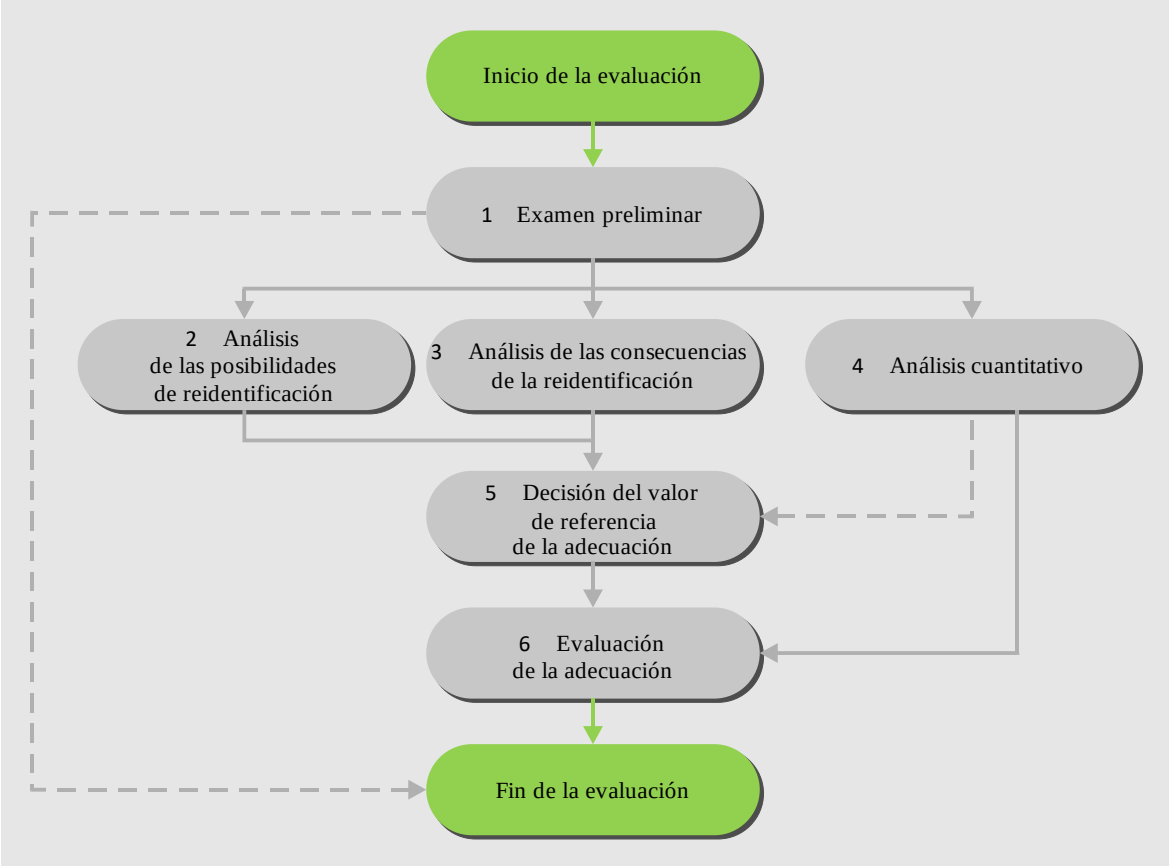



**Figura A.2 – Procedimiento de evaluación de la adecuación**

## A.4 Medidas de desidentificación adicionales

– El controlador de datos aplicará medidas de desidentificación adicionales conforme a lo decidido por el grupo de evaluación, si se considera que la desidentificación no es adecuada.

– El grupo de evaluación procederá a una nueva evaluación cuando el controlador haya terminado de aplicar la desidentificación adicional.

## A.5 Utilización de los datos

– Los datos desidentificados se utilizan para el análisis de macrodatos o se facilitan a terceros si la evaluación (reevaluación) considera que la desidentificación es adecuada.

– En principio, la facilitación o revelación de datos al público o usuarios con los que no se ha concluido un acuerdo está prohibida si no se aplica una estrategia de reducción de riesgos adecuada al modelo de liberación de datos debido al elevado riesgo de reidentificación.

– Los datos se destruirán una vez cumplido el objetivo de utilización o cuando ya no sean necesarios.

– Durante el proceso de utilización de los datos se seguirán las fases de gestión del seguimiento para que la utilización de los datos desidentificados sea efectiva.

# Anexo B

## Métodos de desidentificación no estructurados

(Este anexo forma parte integrante de la presente Recomendación.)

A diferencia de la desidentificación de datos estructurada, que se aplica a campos de datos estructurados, los mecanismos de desidentificación de datos no estructurados se aplica a datos brutos. En el caso de la fotografía siguiente, la desidentificación implica la supresión de las caras o su sustitución por otras, como se muestra en la Figura B.1.

**Figura B.11 – Ejemplo de desidentificación facial**

Hay cuatro tipos de datos no estructurados:

1) datos textuales no estructurados: datos web, informes, blogs, noticias, etc.;
2) datos de vídeo no estructurados: todos los datos de vídeo son no estructurados y algunas etiquetas ofrecen datos regularizados;
3) datos de audio no estructurados: todos los datos de audio son no estructurados y algunas etiquetas o audio reconocido se traducen en datos textuales;
4) datos de registro cronológico no estructurados: los registros cronológicos generados automáticamente no están estructurados, pero suelen seguir un patrón que puede traducirse en forma estructurada.

Para representar la información sintáctica de los datos no estructurados, incluidos el texto, la voz, la imagen y el vídeo, el sistema de desidentificación debe constar de las siguientes tres unidades:

1) Unidad de detección de información multimedios para detectar la información metatextual en los datos multimedios de entrada:
   - incluye un detector de discurso que convierte la voz en texto para rastrear un objeto o actividad incluido en la voz;
   - incluye un detector óptico de reconocimiento de caracteres que extrae caracteres de las imágenes;
   - incluye un detector visual que extrae los objetos o actividades incluidos en una imagen o las imágenes suprimidas de una imagen fija o en movimiento;
   - incluye un detector visual de frases que extrae frases textuales de una imagen fija o en movimiento.

2) Unidad de conformación por conocimiento, que divide la información metatextual y la información contextual en una configuración extrínseca de representación sintáctica e información intrínseca de representación semántica:
   - la información sintáctica comprende la información original generadora de datos multimedios, información de los datos multimedios generados por la fuente, y la información de detección de objetos extraída de una región significativa;
   - la información semántica incluye la información de evento incluida en la región significativa que configura los datos multimedios y la información contextual.
3) Unidad de desidentificación, que suprime la IIP identificable de la base de conocimiento y la información contextual.

Para los datos no estructurados, los mecanismos de desidentificación deben definir los requisitos y niveles de seguridad correspondientes de la siguiente manera:

- Objetivo de la desidentificación: identificar el objetivo que debe protegerse para la aplicación o los servicios en línea.
- Modo de desidentificación: identificar el mecanismo que debe aplicarse para la desidentificación; ¿cuál es el nivel de desidentificación (por ejemplo, Black box, pixelación, difuminación)?
- Desidentificación y reidentificación: hay que determinar si es necesario recuperar o reidentificar los datos. En caso de que la policía necesite una fotografía original para investigar un delito, ¿es posible recuperar la fotografía desidentificada?

# Apéndice I

## Ejemplos de técnicas de desidentificación típicas

(Este apéndice no forma parte integrante de la presente Recomendación.)

Este apéndice contiene ejemplos y descripciones de las técnicas de desidentificación típicas.

### I.1 Herramientas estadísticas para técnicas de desidentificación

– Muestreo: proceso mediante el cual se libera una muestra del conjunto de datos, en lugar de su totalidad. Al liberarse una submuestra, se reducen las probabilidades de reidentificación.

– Agregación: conjunto de funciones estadísticas que producen el valor representado de todo un conjunto de datos.

### I.2 Herramientas criptográficas para técnicas de desidentificación

– Encriptación determinista [b-ISO/CEI 11770]: esquema de encriptación que siempre produce el mismo texto cifrado para un texto no encriptado y una clave dados en distintas ejecuciones del algoritmo de encriptación.

– Encriptación con preservación del orden [b-AGRAWAL]: esquema de encriptación en que se preserva el orden numérico de los textos no encriptados.

– Encriptación homomórfica [b-ISO/CEI 18033-6]: esquema de encriptación que permite efectuar cálculos en el texto cifrado, generando así una encriptación que coincide con el resultado de las operaciones que se efectuarán en el texto no encriptado tras su desencriptación.

– Encriptación con preservación del formato [b-NIST 800-38G]: esquema de encriptación en que el texto cifrado tiene el mismo formato que el texto no encriptado.

– Compartición de secreto homomórfica [b-ISO/CEI 18033-6]: tipo de algoritmo de compartición de secreto en el que el secreto se encripta de manera homomórfica.

### I.3 Técnicas de supresión

– Ocultación: proceso mediante el cual se sustituye un campo con un valor o se elimina. Ejemplos de técnicas de supresión son la sustitución de un número de teléfono por asteriscos o la generación aleatoria de pseudónimos.

– Supresión local: proceso mediante el cual se suprimen o eliminan valores específicos de atributos de los registros seleccionados. La eliminación de los datos aumenta la protección de la privacidad, pero reduce la utilidad de los datos.

– Supresión de registros: proceso que implica la eliminación de uno o varios registros de un conjunto de datos.

### I.4 Técnicas de pseudonimización

Proceso que elimina la asociación con el sujeto de datos y crea una asociación entre una serie concreta de características relativas al sujeto de datos y uno o más pseudónimos. Por norma general la pseudonimización se efectúa sustituyendo identificadores directos por pseudónimos, como valores generados aleatoriamente. Los identificadores directos son, entre otros, los nombres, las direcciones de correo electrónico y los números asignados por el Estado. Se sustituyen por pseudónimos todos los identificadores directos y, posiblemente, algunos o todos los atributos identificadores restantes.

### I.5 Técnicas de generalización

– Redondeo: proceso que consiste en sustituir un valor numérico por otro aproximadamente igual, pero cuya representación es más corta, más simple o más explícita.

– Codificación superior-inferior: proceso mediante el cual los atributos cuyos valores están por encima (o por debajo) de un límite superior (o inferior) dado se definen como umbrales para el mayor (o menor) valor posible.

### I.6 Técnicas de aleatorización

– Adición de ruido: proceso mediante el cual se añade en un atributo dado del conjunto de datos un valor aleatorio imposible de predecir.

– Permutación: proceso mediante el cual se intercambian los valores de un atributo seleccionado en todos los registros del conjunto de datos sin modificarlos.

– Microagregación: proceso mediante el cual se sustituyen todos los valores de atributos continuos por sus medias calculadas de manera algorítmica.

### I.7 Datos sintéticos

Con el método de datos sintéticos se generan artificialmente microdatos para representar un modelo de datos estadístico predefinido. Por definición, los conjuntos de datos sintéticos no contienen datos obtenidos de sujetos de datos existentes, pero parecen reales para el objetivo de utilización previsto.

# Apéndice II

# Métodos de desidentificación

(Este apéndice no forma parte integrante de la presente Recomendación.)

Este apéndice presenta contiene ejemplos y detalles de los métodos de desidentificación.

## II.1 Método de desidentificación centrado en los datos

Dado que las técnicas de desidentificación modifican los datos originales para evitar la revelación de IIP, es evidente que hay que elegir entre utilidad y privacidad. El reto es proteger la privacidad con una pérdida mínima de exactitud. Lo ideal sería que los usuarios de datos analizasen los datos desidentificados sin perder exactitud en relación con los resultados que daría el análisis de los datos originales.

En la práctica resulta difícil realizar una desidentificación perfecta sin poner en peligro la utilidad de los datos. En el caso de los macrodatos este problema se agudiza a causa de la cantidad y variedad de datos. Por una parte, no suele bastar un bajo nivel de desidentificación (por ejemplo, desidentificación mediante la simple supresión de los identificadores directos) para garantizar la ausencia de identificabilidad. Por otra parte, una desidentificación demasiado fuerte puede evitar la vinculación de los datos de una misma persona (o de personas semejantes) procedentes de diversas fuentes y, por tanto, anular muchos posibles beneficios de los macrodatos.

En esta cláusula se describen dos métodos de desidentificación centrada en los datos para equilibrar la utilidad y la privacidad. Las medidas de utilidad genéricas, específicas para el uso de los datos, pueden emplearse para definir cómo medir la utilidad de un conjunto de datos liberado desidentificado.

### II.1.1 Método de desidentificación proutilidad

En los macrodatos, la información sobre una persona suele obtenerse de varias fuentes independientes. Por tanto, la capacidad de vincular registros pertenecientes a la misma persona (o personas del mismo tipo/similares) es esencial a la hora de crear macrodatos.

En el método de desidentificación proutilidad se aplica a un microconjunto de datos una técnica de desidentificación con elección heurística de parámetros y propiedades de preservación de la utilidad adecuadas, tras lo cual se mide el riesgo de revelación. Por consiguiente, el método de desidentificación proutilidad es un proceso lento que carece de garantías de privacidad formales. Por ejemplo, el riesgo de reidentificación puede estimarse empíricamente intentando vincular los datos originales con los desidentificados. Si se considera que el riesgo existente es demasiado elevado, deberá volver a aplicarse la técnica de desidentificación con parámetros de privacidad más estrictos y, probablemente, sacrificar la utilidad, modificando iterativamente los parámetros hasta que el riesgo empírico de revelación sea lo suficientemente bajo, como se suele hacer en las estadísticas oficiales.

Si bien la vinculabilidad es deseable desde el punto de vista de la utilidad, también resulta una amenaza a la privacidad: la precisión de la vinculación debe ser mucho menor en los datos desidentificados que en los originales. El grado de vinculabilidad compatible con una técnica de desidentificación o un modelo de privacidad de desidentificación determina si un analista puede vincular datos desidentificados por separado (con esa técnica/modelo) que corresponden al mismo individuo y cómo puede hacerlo.

### II.1.2 Método de desidentificación proprivacidad

Se aplica un modelo de privacidad con un parámetro que garantiza un límite superior del riesgo de revelación y reidentificación y, quizá, también del riesgo de revelación de atributos. El modelo se aplica con una técnica de desidentificación propia cuyos parámetros se derivan de los del modelo. Entre los modelos de privacidad reconocidos están la anonimia-k y sus extensiones, así como la privacidad diferencial-ε, que suelen ofrecer una baja utilidad/vinculabilidad de los datos.

En el método de desidentificación proprivacidad, si la utilidad de los datos desidentificados resultantes es demasiado baja, el modelo de privacidad se aplicará con una técnica de desidentificación alternativa, menos perjudicial para la utilidad, con un parámetro de privacidad menos estricto o, incluso, podrá recurrirse a un modelo de desidentificación diferente.

## II.2 Método de desidentificación centrado en la función

En esta cláusula se describen tres tipos de métodos que ejecutan las funciones y responsabilidades de cada uno de ellos en el proceso de desidentificación. El método centrado en la función puede caracterizarse en términos generales por la respuesta a las preguntas ¿quién?, ¿qué? y ¿dónde y cómo?:

– ¿Quién tiene acceso a los datos?

– ¿Qué análisis pueden o no pueden hacerse?

– ¿Dónde tiene lugar el acceso/análisis de los datos y cómo se obtiene ese acceso?

### II.2.1 Desidentificación centralizada

El proceso de control de revelación estadístico se fija en la desidentificación centralizada, efectuada por un controlador de datos que tiene acceso a todo el conjunto de datos original. Este método centralizado tiene sus ventajas e inconvenientes, como se muestra en el Cuadro II.1.

**Cuadro II.1 – Características de la desidentificación centralizada**

| | **Detalles** |
|---|---|
| **Ventajas** | • Las personas no necesitan desidentificar los registros de datos que facilitan. Puede esperarse que el controlador de datos, que tiene más recursos de cálculo y probablemente más experiencia en la desidentificación, desidentifique adecuadamente todo el conjunto de datos.<br>• El controlador de datos tiene una visión global del conjunto de datos original y, por tanto, está en la mejor posición para optimizar el equilibrio entre la utilidad de los datos y el riesgo de revelación existente. |
| **Inconvenientes** | • Todas las partes que ofrecen datos originales deben confiar en el controlador de datos (porque tiene acceso a todos los datos originales). Si bien esto no es un problema en las estadísticas oficiales, donde el controlador de datos es un instituto nacional de estadística, puede ser un obstáculo notable en el caso típico de los macrodatos, cuando, por ejemplo, el controlador de datos que reúne varias fuentes de datos es una empresa privada (por ejemplo, un corredor de datos).<br>• En particular en el caso de los macrodatos, la desidentificación puede ser una carga computacional demasiado pesada para un único controlador.<br>• En el procesamiento de macrodatos participan muchos controladores, por lo que el método centralizado es imposible de gestionar. |

Los métodos de desidentificación local y desidentificación colaborativa complementan las ventajas e inconvenientes expuestos.

### II.2.2 Desidentificación local

La desidentificación local es un método de limitación de la revelación alternativo adaptado a casos (incluidos los de macrodatos) en que las personas (sujetos de datos) no confían (o sólo parcialmente) en el controlador de datos que reúne los datos. Cada sujeto desidentifica sus propios datos antes de entregarlos al controlador de datos.

Teniendo en mente la protección de la privacidad, los datos obtenidos por una fuente dada deben desidentificarse en el origen antes de ponerlos a disposición. Sin embargo, la desidentificación independiente aplicada por cada fuente genera una pérdida de información superior a la de la desidentificación centralizada, porque los sujetos desidentifican sus datos sin ver los datos de otros sujetos. Quiere esto decir que los sujetos carecen de una visión global del conjunto de datos, por lo que les resulta difícil hallar el equilibrio entre la limitación del riesgo de revelación y la pérdida de información.

### II.2.3 Desidentificación colaborativa

La desidentificación colaborativa combina la escasa pérdida de utilidad de la desidentificación centralizada y la elevada privacidad de la desidentificación local. Uno de los problemas de la desidentificación centralizada es que, si el sujeto de datos no confía en que el controlador de datos va a utilizar y/o desidentificar adecuadamente sus datos, puede decidir facilitar datos falsos (creando así un sesgo en la respuesta) o no facilitar datos en absoluto (creando así un sesgo de no respuesta). Por consiguiente, los sujetos pueden colaborar para determinar el riesgo de revelación asociado a sus datos y aplicar localmente el nivel adecuado de protección de manera distribuida y colaborativa, cuyas principales propiedades son las siguientes:

– No se pierde más información del conjunto de datos de la que se perdería con el método centralizado con el mismo nivel de privacidad. Es mejor que el método local, pues se pierde menos información.

– Ni los sujetos de datos ni el controlador de datos pueden conocer más atributos confidenciales de otro sujeto de datos dado que los contenidos en el conjunto de datos desidentificado final. Es mejor que el método centralizado, pues ofrece también privacidad con respecto al colector de datos.

Además, el método colaborativo puede llevar a protocolos que funcionen bien sin mecanismos de aplicación externos. En la desidentificación de microdatos, la protección de la privacidad que obtiene el sujeto afecta a la protección de la privacidad de otros implicados. Para mejorar la coutilidad en el método colaborativo, es necesaria la transformación multipartita segura para que los protocolos electrónicos permitan a dos o más partes llevar a cabo una transformación que implica a sus conjuntos de datos de manera que ninguna de las partes tenga que entregar explícitamente un conjunto de datos a cualquiera de las otras partes. Al permitir la transformación de las búsquedas sin necesidad de centralizar el almacenamiento de los datos, la transformación multipartita segura reduce los daños causados por una fuga de datos y permite efectuar cálculos entre partes que no confían plenamente unas en otras. La computación multipartita puede ofrecer al tiempo una mejor privacidad y una mayor utilidad en determinados contextos.

# Bibliografía


[b-ISO/IEC 11770] ISO/IEC 11770 (all parts), *Information technology – Security techniques – Key management*.

[b-ISO/IEC 18033-6] ISO/IEC 18033-6, *Information technology security techniques – Encryption algorithms – Part 6: Homomorphic encryption.*

[b-ISO/IEC 20889] ISO/IEC 20889 (2018), *Privacy enhancing data de-identification terminology and classification of techniques.*

[b-ISO/IEC 27001] ISO/IEC 27001 (2018), *Information technology – Security technique – Information security management systems.*

[b-ISO/IEC 29100] ISO/IEC 29100 (2011), *Information technology – Security technique – Privacy framework.*

[b-NIST 800-38G] NIST Special Publication 800-38G (2016), *Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption.*

[b-NISTIR 8053] NISTIR 8053 (2015), *De-Identification of Personal Information.*

[b-AGRAWAL] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2004), *Order preserving encryption for numeric data, SIGMOD '04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data, Paris, France, June, pp 563-574.*

[b-KOREA] Korean Ministry of the Interior, *Guidelines on De-identification Measures, June 2016.*
<http://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000821178&fileSn=2&nttId=7187&toolVer=&toolCntKey_>
Last accessed 26 July 2019)
<https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000827161&fileSn=0>
(English, last accessed 12 December 2020)

[b-UKAN] UK Anonymization Network, *The anonymisation decision-making framework, 2016*
<https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>

## SERIES DE RECOMENDACIONES DEL UIT-T

| | |
|---|---|
| Serie A | Organización del trabajo del UIT-T |
| Serie D | Principios de tarificación y contabilidad y cuestiones económicas y políticas de las telecomunicaciones/TIC internacionales |
| Serie E | Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos |
| Serie F | Servicios de telecomunicación no telefónicos |
| Serie G | Sistemas y medios de transmisión, sistemas y redes digitales |
| Serie H | Sistemas audiovisuales y multimedia |
| Serie I | Red digital de servicios integrados |
| Serie J | Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedia |
| Serie K | Protección contra las interferencias |
| Serie L | Medio ambiente y TIC, cambio climático, ciberdesechos, eficiencia energética, construcción, instalación y protección de los cables y demás elementos de planta exterior |
| Serie M | Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes |
| Serie N | Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión |
| Serie O | Especificaciones de los aparatos de medida |
| Serie P | Calidad de la transmisión telefónica, instalaciones telefónicas y redes de líneas locales |
| Serie Q | Conmutación y señalización, y mediciones y pruebas asociadas |
| Serie R | Transmisión telegráfica |
| Serie S | Equipos terminales para servicios de telegrafía |
| Serie T | Terminales para servicios de telemática |
| Serie U | Conmutación telegráfica |
| Serie V | Comunicación de datos por la red telefónica |
| **Serie X** | **Redes de datos, comunicaciones de sistemas abiertos y seguridad** |
| Serie Y | Infraestructura mundial de la información, aspectos del protocolo Internet, redes de próxima generación, Internet de las cosas y ciudades inteligentes |
| Serie Z | Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación |

Impreso en Suiza
Ginebra, 2021