COMISIÓN PARA EL MERCADO FINANCIERO

Regulador y Supervisor Financiero de Chile

Nota Técnica **N° 02/24**

Anonimización de datos de deudores en instituciones financieras

Jaime Forteza – Sebastián Ramírez

Octubre 2024
www.CMFChile.cl

COMISIÓN PARA EL MERCADO FINANCIERO

Regulador y Supervisor Financiero de Chile

The Technical Notes Series is a publication of the Financial Market Commission (CMF), whose purpose is to contribute with short articles to the discussion of issues relevant to financial stability and financial regulation. Although these notes have the editorial revision of the CMF, the analysis and conclusions set forth are the responsibility of the authors and do not necessarily reflect the views of the CMF.

La serie de Notas Técnicas es una publicación de la Comisión para el Mercado Financiero (CMF), cuyo objetivo es aportar con artículos breves al debate de temas relevantes para estabilidad y regulación financieras. Si bien estas notas cuentan con la revisión editorial de la CMF, los análisis y conclusiones en ellos contenidos son de exclusiva responsabilidad de sus autores.

Nota Técnica de la Comisión para el Mercado Financiero (CMF) Financial Market Commission (CMF)
Av. Libertador Bernardo O'Higgins 1449, Santiago, Chile Teléfono: (56) 22617 4058

Nota Técnica N° 01/24
Technical Note N° 01/24

Anonimización de datos de deudores en instituciones financieras.*

Jaime Forteza Saavedra[1], Sebastián Ramírez Venegas [2]

Octubre 2024

RESUMEN

A nivel internacional, diversas jurisdicciones han adoptado técnicas de anonimización de datos para permitir el tratamiento de información histórica y la vez, cumplir con las leyes de protección de datos personales. A pesar del riesgo potencial de desanonimización o reidentificación, se han desarrollado diversos métodos de anonimización que ayudan a mitigar este riesgo. En el contexto bancario chileno, uno de los objetivos de los procesos de anonimización es facilitar el uso de modelos internos por parte de la industria financiera. Esto permite realizar estimaciones considerando un ciclo económico completo, superando así la restricción legal actual de 5 años para datos personales relacionados con obligaciones financieras, impuestas por la Ley sobre Protección de la Vida Privada. Esta nota técnica busca establecer un marco metodológico general basado en la experiencia internacional, junto con proporcionar orientación a las instituciones financieras sobre las técnicas de anonimización más utilizadas y ofrecer lineamientos técnicos para el tratamiento adecuado de estos tipos de datos. Se presentan ejemplos para implementar procesos de disociación que se adapten mejor a las características específicas de los datos que requieran ser anonimizados, identificando los posibles riesgos de reidentificación de cada técnica.

ABSTRACT

Internationally, various jurisdictions have implemented anonymization techniques to facilitate the processing of historical personal data while ensuring compliance with data protection laws. Despite the inherent risk of de-anonymization or re-identification, contemporary legislative frameworks often proactively address these challenges to enable the secure and lawful utilization of such data. In the Chilean banking sector, anonymization plays a pivotal role in supporting the development of internal models by financial institutions. By anonymizing data, institutions can extend their analytical horizon beyond the statutory limitation of 5 years imposed by the Law on Protection of Private Life, thereby enabling more comprehensive economic cycle estimations. This technical note endeavors to establish a robust methodological framework based on international best practices. Its primary objective is to provide financial institutions with guidance on the adoption of widely recognized anonymization techniques and to delineate precise methodologies for the proper handling of data. Special emphasis is placed on implementing dissociation processes tailored to the specific characteristics of the data requiring anonymization, while maintaining a keen awareness of the potential re-identification risks associated with each technique.

*/ Las opiniones emitidas en este trabajo, errores y omisiones, son de exclusiva responsabilidad de los autores y no necesariamente reflejan la visión de la institución. Se agradecen los comentarios, consejos y sugerencias del referato interno, así como también los de otros participantes en seminarios internos.

[1]/ División Normativa de Regulación Prudencial, Dirección de Regulación de Bancos e Instituciones Financieras, Dirección General de Regulación Prudencial, CMF, jforteza@cmfchile.cl

[2]/ División Normativa de Regulación Prudencial, Dirección de Regulación de Bancos e Instituciones Financieras, Dirección General de Regulación Prudencial, CMF, sramirez@cmfchile.cl

I. Introducción

Técnicamente, es factible establecer un sistema de información de deudores donde todos los datos derivados de los procesos de solicitud de créditos, historial de pagos, insolvencia, entre otros, sean almacenados, clasificados y procesados para su uso en la intermediación financiera, permitiendo mitigar los problemas de asimetría de información entre prestamistas y prestatarios. No obstante, además de las limitaciones tecnológicas para mantener dicho sistema, existen restricciones legales que rigen la preservación de esta información. De esta forma, los sistemas de datos del sector financiero deben cumplir rigurosamente con las leyes y regulaciones que protegen la privacidad, los datos personales y el derecho al olvido, asuntos particularmente sensibles, así como también lo son las normativas sobre secreto bancario y reserva.

Desde 2013, la Asamblea General de las Naciones Unidas y su Consejo de Derechos Humanos han aprobado numerosas resoluciones sobre el derecho a la privacidad en la era digital. Estas directrices se han materializado en la mayoría de las jurisdicciones a nivel mundial, a través de leyes que regulan el manejo de la información sensible del cliente financiero, estableciendo métodos adecuados para su protección. En Chile, la Ley N° 19.628 sobre Protección de la Vida Privada define como datos personales toda información concerniente a personas naturales, identificadas o identificables. Específicamente, el artículo 18 de esta ley establece que los datos personales relacionados con obligaciones de carácter económico, financiero, bancario o comercial no pueden ser comunicados[3], si se refieren a una persona identificada o identificable, después de transcurridos cinco años desde que la obligación respectiva se hizo exigible, o después de haber sido pagada o extinguida de otra manera legal.

Una solución adoptada por diversas jurisdicciones para utilizar información caduca y cumplir al mismo tiempo con las leyes de protección, es la anonimización o disociación de los datos personales (Garrido et al., 2019). Este proceso tiene como objetivo ocultar o eliminar cualquier vínculo entre un conjunto de datos y la identificación de sus titulares. La anonimización requiere comprender adecuadamente el objetivo final para el cual se necesita utilizar la información y evaluar su utilidad para dicho propósito. Independiente de las técnicas de anonimización empleadas, una vez que se ha realizado la disociación, la información original del conjunto de datos se ve reducida. Sin embargo, en la práctica, la anonimización no siempre previene por completo la conexión de los titulares son su información debido al riesgo de desanonimización o reidentificación, es decir, la posibilidad de volver a vincular la identidad con los datos (Ohm, 2009). Las legislaciones actuales no pueden pasar por alto el riesgo de reidentificación, lo que obliga a los distintos reguladores a responder de manera rápida y efectiva a este desafío tecnológico para proteger a los titulares de la información contra posibles daños significativos (Brasher, 2018; Ohm, 2009).

En este contexto, las leyes de privacidad modernas adoptan un enfoque preventivo que permite el tratamiento de información histórica de larga data, buscando combinar los beneficios del flujo de información con sólidas garantías de privacidad. Para cumplir con las leyes de protección de datos personales, las normativas internacionales se centran en gestionar los riesgos de reidentificación en lugar de buscar una anonimización perfecta. Esto incluye implementar políticas robustas de divulgación y tratamiento de datos personales por parte de las entidades, donde los procedimientos deben considerar estudios sobre los riesgos que se pueden presentar si los datos consiguen vincularse nuevamente con la información anonimizada, y, por ello, se consideran todas las posibles consecuencias que se pueden dar en caso de que se reidentifique información sensible.

3 Dar a conocer de cualquier forma los datos de carácter personal a personas distintas del titular, sean determinadas o indeterminadas.

Regulaciones como el Gramm-Leach-Bliley Act (GLBA) en EE. UU y General Data Protection Regulation (GDPR) de la Unión Europea obligan a los bancos a auditar sus datos trimestralmente y eliminar los datos personales caducos, a menos que, cuenten con el consentimiento explícito de los clientes. No obstante, para permitir a los bancos seguir utilizando esta valiosa información para otros fines distintos al otorgamiento de créditos, se les permite anonimizarla. Particularmente el GDPR considera como una infracción muy grave la reversión deliberada del proceso de anonimización con el fin de permitir la reidentificación de los titulares, estableciendo umbrales de tolerancia a este riesgo y proporcionando garantías adicionales para proteger la privacidad de los clientes. Este reglamento impone estrictas sanciones y multas administrativas a quienes violen los estándares de privacidad y seguridad, las cuales pueden ascender hasta 20.000.000 EUR o, en el caso de una empresa, hasta el 4% del volumen total de negocios anual global del ejercicio financiero anterior, aplicándose la cuantía más elevada de ambas opciones.

En Chile, la Comisión para el Mercado Financiero (CMF) publicó en julio de 2022 una norma que exige a las instituciones financieras una Política Interna de Seguridad y Manejo de la Información de Deudores[4]. Esta política requiere que las instituciones realicen procedimientos de disociación de datos, eliminando cualquier vínculo con los titulares de la información caduca, especialmente para su uso estadístico o en modelos de riesgo.

Asimismo, durante los últimos años se han impulsado y aprobado diversas iniciativas legales para fortalecer la divulgación de información de deuda, como la reciente creación del Registro de Deuda Consolidada (REDEC). Esta medida está alineada con experiencias de distintas jurisdicciones y recomendaciones de organismos internacionales[5]. En el mismo contexto, se aprobó la Ley FINTEC, la cual establece un marco regulatorio flexible para empresas que ofrecen servicios financieros y promueve el desarrollo de un Sistema de Finanzas Abiertas reguladas, que permitirá a las personas compartir su información financiera para acceder a mejores productos y servicios. Además, actualmente se encuentra en trámite legislativo el Boletín 11.144-07, que regula la protección y tratamiento de los datos personales y crea la Agencia de Protección de Datos Personales en Chile. Estas medidas reflejan un esfuerzo y compromiso de Chile por modernizar su marco regulatorio financiero y de protección de datos, alineándose con estándares internacionales y promoviendo un entorno más transparente y seguro para los usuarios de servicios financieros.

En línea con lo anterior, la anonimización de datos personales desempeña un papel relevante en relación con el uso de modelos internos por parte de la industria financiera. Particularmente, el Capítulo B-1 del Compendio Normas Contables Bancos y el Capítulo 21-6 de la Recopilación Actualizada de Normas (RAN), indican que las instituciones financieras podrán optar por constituir provisiones y capital utilizando metodologías internas, respectivamente. Para aplicar estas metodologías, es requisito contar con al menos 5 años de información histórica, que incluya al menos un período recesivo en la estimación de la probabilidad de incumplimiento (PI) y la pérdida dado el incumplimiento (PDI). La anonimización de datos personales de deudores permite cumplir con estos requisitos mientras se adhiere a los lineamientos de la Ley N° 19.628 sobre Protección de la Vida Privada. Esto asegura que las estimaciones financieras y de riesgo se realicen de manera precisa y legalmente conforme, incluso considerando ciclos económicos completos que superen los 5 años requeridos.

4 Numeral del 6 Capítulo 18-5 de la Recopilación Actualizada de Normas de Bancos sobre Información sobre deudores de las instituciones financieras. Fuente: https://www.cmfchile.cl/portal/principal/613/articles-28970_doc_pdf.pdf

5 Recientemente, el Fondo Monetario Internacional y el Banco Mundial evaluaron el sector financiero chileno recomendando ampliar la cobertura del sistema de información crediticia para incluir datos positivos y abstenerse de eliminar la información crediticia negativa.

Con el objeto de definir lineamientos para desarrollar un proceso estandarizado de control de divulgación de datos, en 2019, el Instituto Nacional de Estadísticas (INE) de Chile publicó la primera versión de la "Guía para el control de divulgación estadística en microdatos[6]", basada en una guía práctica del Banco Mundial (Benschop, Machingauta, & Welch, 2019). Esta guía, actualizada en 2021, proporciona directrices generales para el manejo seguro de datos en base a un proceso estandarizado de doce pasos para controlar la divulgación de microdatos estadísticos, asegurando la seguridad de la información recolectada y procesada por la institución, por lo que su alcance se limita a proveer directrices circunscritas al campo de los microdatos en términos generales.

Esta nota técnica tiene como objetivo establecer un marco metodológico general, principios y protocolos mínimos para el proceso de anonimización de datos personales en la industria financiera, basándose en experiencias internacionales. A su vez, pretende orientar sobre las técnicas más utilizadas de anonimización y proporcionar lineamientos técnicos para el adecuado aprovechamiento y uso de estos datos. El proceso de anonimización debe adaptarse a las características específicas de los datos que se necesitan anonimizar, considerando cuidadosamente los riesgos potenciales de reidentificación. Así, se quiere contribuir a que las instituciones financieras den cumplimiento a la normativa existente en materia de protección de datos personales y, al mismo tiempo, logren maximizan el valor estratégico de esta información como un activo clave, mejorando la eficiencia del sistema en su conjunto.

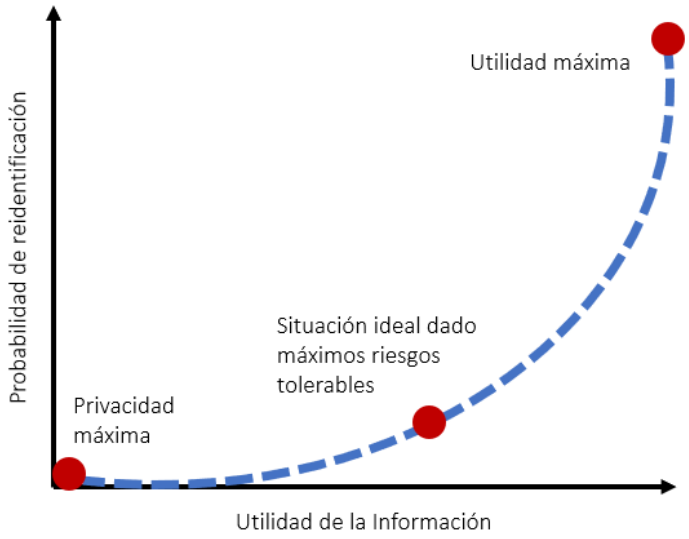
II. Anonimización de datos personales

La anonimización es el proceso mediante el cual se transforma la información con el fin de condicionar un conjunto de datos de tal manera que se eliminen o reduzcan al mínimo las posibilidades de identificación de los individuos. Este proceso se realiza para proteger la privacidad y confidencialidad de los titulares de datos (Data Protection Commission, 2019). Al anonimizar los datos, se permite que la información pueda ser utilizada para llevar a cabo análisis técnico y científico válido sobre el conjunto de datos anonimizado, al mismo tiempo que se protege la privacidad de los titulares (Lasko, 2009). Para lograr este propósito, es fundamental eliminar suficiente información de modo que no se pueda inferir una identidad basada en la información restante. Esta eliminación debe abarcar tanto datos directos, como el nombre, la dirección y los números de teléfono, como datos indirectos que podrían identificar a una persona mediante el cruce de datos anonimizados con otras fuentes de información externas.

Existe un potencial *trade-off* entre el nivel de anonimización de los conjuntos de datos y su utilidad. Como se ilustra en la Figura 1, a medida que se conserva más información no anonimizada, aumenta su utilidad para el análisis técnico, pero también crece la probabilidad de reidentificación de los titulares. Por lo tanto, el proceso de anonimización y cómo ésta se implementa, influye directamente en el riesgo de reidentificación. Un proceso robusto en esta materia debiese reducir al máximo este riesgo, considerando tanto los atributos de la información que se requiere disociar como los controles de mitigación utilizados en el proceso, y la existencia de un riesgo residual de reidentificación. En otras palabras, es necesario determinar un nivel tolerable de riesgo que sea coherente con un nivel óptimo de privacidad de la información, de manera que se permita realizar un análisis con el conjunto de datos anonimizados que no difiera significativamente del análisis que se obtendría con los datos originales, gestionando y monitoreando los riesgos a lo largo de todo el proceso.

6 https://www.ine.gob.cl/docs/default-source/buenas-practicas/estandares/estandar/documento/gu%C3%ADa-control-divulgaci%C3%B3n-estad%C3%ADstica-microdatos.pdf?sfvrsn=fb568638_2

Figura 1. *Trade off* entre el riesgo de reidentificación y la utilidad de la información.



En la práctica, existen diversos métodos para anonimizar datos personales. Para elegir los métodos adecuados, se debe seguir un proceso iterativo en el cual se aplican los métodos seleccionados, y luego, se vuelven a medir los riesgos de reidentificación y la utilidad del nuevo conjunto de datos. Estos resultados se comparan con otros métodos de anonimización y con diferentes configuraciones de parámetros.

Si los resultados son satisfactorios y cumplen con los criterios establecidos, los datos pueden ser utilizados. Sin embargo, si el riesgo de reidentificación no puede ser reducido lo suficiente o si la pérdida de información es demasiado alta, entonces el proceso debe repetirse con métodos alternativos o ajustes en los parámetros. De esta forma, se debe encontrar una solución que satisfaga tanto la minimización del riesgo de identificación como la maximización de la utilidad de los datos para su análisis. Este enfoque iterativo y comparativo es fundamental para equilibrar adecuadamente los objetivos de privacidad y utilidad en el proceso de anonimización de datos.

III. Principios y protocolos para la anonimización de datos personales

En el desarrollo de la anonimización de datos se requiere establecer protocolos y procedimientos específicos para generar una protección de la privacidad desde el diseño y, por defecto, abarcando la implementación y evaluación continua a lo largo de todo el ciclo de vida del sistema de información. En esta sección del documento se abordan, en primer lugar, los principios generales mínimos que se deben considerar para un diseño óptimo de anonimización de datos. Posteriormente, se propone un marco general sobre los protocolos y políticas internas que deberían adoptar las entidades financieras para asegurar un proceso de anonimización efectivo.

A) Principios generales

En la anonimización de datos personales, se necesita que las entidades establezcan sólidos procesos internos de protección de dichos datos desde el diseño, con el objetivo específico de asegurar el uso correcto y el resguardo adecuado de la información. La anonimización debe ser una práctica constante y continua, que cumpla con todas las medidas de protección de los datos, anticipando potenciales pérdidas de privacidad de la información antes de que estas ocurran y produzcan perjuicios a sus titulares.

En este sentido, se debe evitar recabar datos de manera indiscriminada, ya que esto aumenta la posibilidad de comprometer la privacidad de los clientes, por lo que, el proceso de anonimización debe limitarse y considerar un conjunto de datos estrictamente necesarios para la finalidad que se quiere conseguir.

En *El Reglamento General de Protección de Datos* (RGPD) de la Unión Europea se definen distintos principios generales que las instituciones deben cumplir en sus protocolos y procedimientos internos. Estos corresponden a los principios[7] de: a) proactividad, b) privacidad por defecto, c) privacidad objetiva, d) plena funcionalidad, e) privacidad en el ciclo de vida de la información y f) Información y formación, los que se definen a continuación:

a) Proactividad: En el proceso de anonimización deben aplicarse todas las medidas necesarias para garantizar la protección de la privacidad, realizando y revisando los procesos de gestión de riesgos de forma preventiva desde el comienzo del diseño del procedimiento y no como reacción a problemas existentes en el proceso. Para esto, al iniciar un proceso de anonimización, se debe clasificar la información estableciendo una escala según la sensibilidad de los datos, la cual deberá ser conocida por todo el personal implicado en el proceso de anonimización.

b) Privacidad por defecto: Al diseñar un sistema de información es esencial garantizar la confidencialidad. Por lo tanto, desde el inicio del proceso se debe salvaguardar la privacidad teniendo en cuenta la granularidad o grado de detalle final que deben tener los datos anonimizados, es decir, este principio se encarga de que solo se procesen los datos personales necesarios de acuerdo con la finalidad del tratamiento. En este sentido, mantener una escala como la referida en el punto anterior es una herramienta de utilidad para la eliminación de las variables redundante según el objetivo del tratamiento de los datos, atendiendo a diferentes criterios de granularidad preestablecidos por la entidad.

c) Privacidad objetiva: En todo proceso de anonimización existirá un riesgo residual de reidentificación, el cual debe ser asumido y mitigado por el coordinador del proceso de anonimización de la institución, el responsable de protección de datos y el responsable del tratamiento de la información. Para ello, se debe considerar el umbral de riesgo máximo aceptable definido internamente por la institución, el cual dependerá del tipo de dato que se va a anonimizar. Este umbral debe ser comunicado e informado a los distintos destinatarios y usuarios de los conjuntos

[7] Estos principios considerados como parte de las buenas prácticas a nivel internacional también han sido adoptados y adaptados por distintas jurisdicciones, como lo son en el caso latinoamericano el de Colombia (Guía de Anonimización de Datos Estructurados del Archivo General de la Nación Colombia de 2020) y el de Uruguay (Guía sobre Anonimización de Datos la Agencia de Gobierno Electrónico y Sociedad de la Información y del Conocimiento de Uruguay de 2020).

datos anonimizados con motivo de mantener el nivel de riesgo por debajo del umbral deseado y así minimizar la probabilidad de reidentificación.

d) Plena funcionalidad: Para asegurar que los datos anonimizados serán útiles debe tenerse en consideración la finalidad del tratamiento de los datos anonimizados, de forma que, debido al proceso de disociación, no existan distorsiones relevantes en el análisis de la información en relación con el tratamiento que se hubiera realizado con los datos no anonimizados.

e) Privacidad en el ciclo de vida de la información: Las medidas que garantizan la privacidad de los titulares de la información son aplicables durante todo el ciclo de vida de la información, partiendo siempre desde la información nominada. Luego, en el proceso de anonimización se eliminarán todos los datos identificativos que no se consideren necesarios para el tratamiento de la información. Finalmente, para el sistema de información anonimizada, se deberán realizar auditorías continuas en las que se verifique el correcto tratamiento de la información por parte de la entidad y de todo el personal relacionado en el proceso de disociación.

f) Información y formación: El personal encargado, tanto de realizar el proceso de anonimización como de tratar esta información, debe tener la adecuada información y formación sobre sus obligaciones. Lo anterior debe contemplarse desde el diseño conceptual del sistema de información y durante todo su desarrollo, teniendo en cuenta los diferentes perfiles de acceso y necesidades de cada uno de los actores involucrados en el proceso de anonimización.

B) Protocolos y políticas para la anonimización.

Es altamente recomendable que las instituciones definan una política interna y protocolos de anonimización, que incluyan un esquema claro del flujo de actividades pertinentes y sus etapas, de modo de orientar al personal que utilice información de datos personales.

Las etapas del proceso de anonimización que se detallan a continuación, se basan en estándares generales de diferentes jurisdicciones que pueden servir de orientación a la hora de definir un protocolo, pero en ningún caso, deben tomarse como un marco o esquema conceptual cerrado. Estas son:

a) Definición de los equipos de trabajo

En el desarrollo de un proceso de anonimización intervienen distintas funciones, las cuales se pueden establecer en base a distintos roles o perfiles de acceso. Dado esto, es altamente conveniente que los procesos de anonimización mantengan claramente definidas las funciones necesarias y el detalle del alcance de cada uno de los perfiles a los cuales pudieran designarse estas funciones. Esto debe realizarse garantizando, dentro de lo posible, que cada uno de los participantes tenga independencia del resto, con el fin de evitar que un error que se produzca a un determinado nivel sea supervisado y aprobado por la misma persona, pero en un nivel diferente. Todos quienes tengan acceso a este conjunto de información siendo parte de los equipos de trabajo durante cualquier etapa del proceso, deberán guardar secreto o confidencialidad acerca del mismo, recomendándose que existan cláusulas de confidencialidad y seguridad de los datos en los contratos de los trabajadores que cumplan estas funciones relacionadas.

En este sentido, dentro de los distintos equipos de trabajo se deberían tener en cuenta al menos los siguientes perfiles:

i. Responsables de protección de datos: debe proporcionar asesoramiento y recomendaciones a la organización y a su personal en relación con las obligaciones de protección de datos. Esto incluye la realización de evaluaciones de impacto previas, verificar la ejecución y velar por la independencia de roles y funciones en todos los procesos de protección de datos de la organización. Debe promover auditorías internas y externas para asegurar el cumplimiento de las políticas de datos y las leyes de protección de datos aplicables. En cuanto a la anonimización de datos personales, debe evaluar periódicamente los riesgos asociados con las actividades de procesamiento de datos, las técnicas de anonimización utilizadas, los costos de implementación, el alcance, el contexto y los fines del tratamiento definidos previamente, así como los riesgos de reidentificación que puedan afectar a los titulares de los datos. Además, como parte de sus responsabilidades debe garantizar la implementación de medidas técnicas apropiadas para asegurar un nivel de seguridad adecuado al riesgo. En caso de que se produzca una violación de la seguridad que implique la filtración, pérdida o alteración accidental o ilícita de datos personales, debe informar de inmediato cuando exista un riesgo razonable para los derechos y libertades de los titulares de los datos. El responsable de protección de datos deberá reportar directamente al gerente de la institución.

ii. Líder/coordinador del proceso de anonimización: persona encargada de decidir la finalidad principal del uso de los datos anonimizados según los objetivos a los que responde esta información. Además, esta persona estará encargada de liderar, orientar y de establecer la conexión dentro de todo el equipo del proceso de anonimización, además de actuar como vínculo con otros equipos/instancias de la entidad.
Deberá ser responsabilidad del directorio de la institución o la máxima autoridad de la organización, la designación al líder del proceso de anonimización. No obstante, este líder deberá contar con autonomía respecto de la administración de la institución. El líder del proceso de anonimización puede desempeñar otras funciones y cometidos dentro de la entidad, siempre que dichas funciones y cometidos no den lugar a conflicto de intereses. El coordinador del proceso de anonimización deberá reportar directamente al responsable de protección de datos de la institución.

iii. Destinatario de la información anonimizada: usuarios finales del conjunto de información anonimizada quienes deben tratar los datos exclusivamente en base a los objetivos finales planteados anteriormente por la entidad.

iv. Equipo de evaluación de riesgos: Para este rol estarán los encargados de realizar la evaluación de riesgos iniciales y los resultados finales del proceso de anonimización, auditar internamente el procedimiento de anonimización y el uso que se le dé a la información anonimizada. Son los responsables de última instancia en asegurar que la información anonimizada cumpla con los requisitos relativos al riesgo residual de reidentificación definido por la entidad. Para procesos no estandarizados, los miembros de este equipo requieren la presencia de analistas temáticos que tengan cabal conocimiento sobre el fenómeno medido, necesidades de información de los usuarios, propósito de la operación estadística o del registro administrativo, cobertura de los datos, indicadores y vinculación con datos anteriores y, fuentes de información externa que pueda ser utilizada por un intruso para la reidentificación de una persona.

v. Equipo de preanonimización y de anonimización: Este equipo se encargará de identificar las variables que deben anonimizar y de proponer y aplicar las técnicas de anonimización necesarias dado los atributos de la información, las cuales deben ser previamente validadas por el equipo evaluador de riesgo.

vi. Equipo de informática: Equipo encargado de la manipulación y procesamiento de los datos mediante el uso de herramientas de software para la anonimización de las bases de datos. Dado que el desarrollo de técnicas de anonimización para procesos no estandarizados suele ser iterativo, la presencia de encargados y analistas temáticos en este proceso es fundamental para apoyar al área de informática.

vii. Equipo de seguridad de la información del proceso de anonimización: este equipo será el encargado de velar por las medidas de seguridad que pudieran ser necesarias durante todo el ciclo de vida de la información anonimizada.

b) Definición de objetivos y finalidad de la información anonimizada

El diseño del proceso de anonimización de datos está definido en base a la finalidad del tratamiento de la información anonimizada, donde la utilización de la información puede ser considerada de datos abiertos o de uso restringido. En este último caso, el tratamiento de datos y el proceso de anonimización puede estar acompañados de acuerdos de confidencialidad con cláusulas específicas sobre protección ante la reidentificación y garantías de la privacidad de la información, las cuales formarán parte del conjunto de las garantías jurídicas en caso de que se determine la identidad de algún titular.

En el contexto de datos de deudores de instituciones financieras, un diseño con datos de uso restringido cobra alta relevancia debido a las restricciones legales que posee este tipo de información. Particularmente, la Ley N°19.628 sobre "Protección de la Vida Privada", en su artículo 18, establece que: "En ningún caso pueden comunicarse los datos a que se refiere el artículo anterior, que se relacionen con una persona identificada o identificable, luego de transcurridos cinco años desde que la respectiva obligación se hizo exigible". Por otro lado, los estándares de Basilea e IFRS 9 exigen un ciclo económico completo de observación para la estimación de los parámetros de la probabilidad de incumplimiento y la perdida dado el incumplimiento. La literatura identifica que la duración promedio de los ciclos económicos varía entre países, donde los países emergentes presentan una duración promedio del ciclo en torno a 10 años, siendo estas más dispersa respecto a la de los países desarrollados. En particular, se estima que para Chile la duración promedio de un ciclo económico es de 8 años (Rand y Tarp, 2002). En este contexto, dado que la duración promedio de un ciclo económico suele superar los cinco años establecidos por la ley para la conservación de datos innominados, es crucial diseñar el proceso de anonimización y los acuerdos de confidencialidad con cuidado. Estos mecanismos no solo aseguran el cumplimiento de los requisitos legales y reglamentarios, sino que también garantizan la protección de la información, permitiendo su uso seguro y conforme a las normativas vigentes.

c) Evaluación de riesgos de reidentificación

Uno de los mayores riesgos de la anonimización de una base de datos proviene de la probabilidad de reidentificar a un individuo dentro de un conjunto. Como se recalcó en el inicio del documento, es importante tener en cuenta que ninguna técnica de anonimización podrá garantizar efectividad absoluta, ya que, existirá siempre una probabilidad de reidentificación que se debe intentar reducir al máximo mediante la correspondiente gestión de riesgos tolerables.

Algunos de los posibles riesgos para la reidentificación de personas con datos anonimizados pueden originarse por: i) la inadecuada implementación de los procedimientos de anonimización y/o tratamiento de estos datos ii) falta de formación y capacitación permanente del personal implicado iii) inadecuada gestión de las llaves/claves utilizadas; entre otros factores.

El responsable del proceso debe tener en consideración la variación evolutiva de los riesgos a lo largo del tiempo. Por ello, es recomendable la reevaluación periódica del riesgo residual existente con el fin de introducir parámetros de mejora de la calidad del proceso de anonimización. Dado lo anterior, es conveniente realizar un análisis de riesgos del proceso de anonimización para posteriormente poder gestionar los riesgos resultantes con medidas técnicas, administrativas o de cualquier otra índole.

Para la medición del riesgo de reidentificación es necesario establecer los umbrales para las distintas fuentes de información que se busca anonimizar, y determinar si el riesgo es alto, medio o bajo. Para ello, generalmente se utilizan modelos en el que se establecen los riesgos y sus probabilidades de ocurrencia, identificando el potencial daño que causaría a los titulares de los datos en caso de que ocurra algún tipo de reidentificación, así como las consecuencias económicas, legales y reputacionales para la entidad que los maneja[8].

Las Guías de anonimización para datos estructurados de Canadá[9], Colombia[10] y España[11] contemplan una medición del riesgo de carácter cuantitativo que incorpora a su vez insumos cualitativos, describiendo ampliamente las posibilidades y los beneficios que proporciona una correcta evaluación del impacto en la protección de los datos personales para los procesos y tratamientos de datos anonimizados.

8 En 2021 la Agencia Española de Protección de Datos publicó una guía enfocada en la gestión del riesgo de los tratamientos de datos personales y evaluaciones de impacto. El documento está dirigido a personas responsables, encargadas de tratamientos y a personas delegadas de protección de datos, ayudando al cumplimiento de la normativa europea. No obstante, esta guía sirve como pauta, siendo aplicable a cualquier proceso de medición de riesgos para el tratamiento de datos personales. Fuente: https://www.aepd.es/es/documento/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf

9 Fuente: https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

10 Fuente: https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/Guia_de_Anonimizacion-min.pdf

11 Fuente: https://www.aepd.es/es/documento/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf

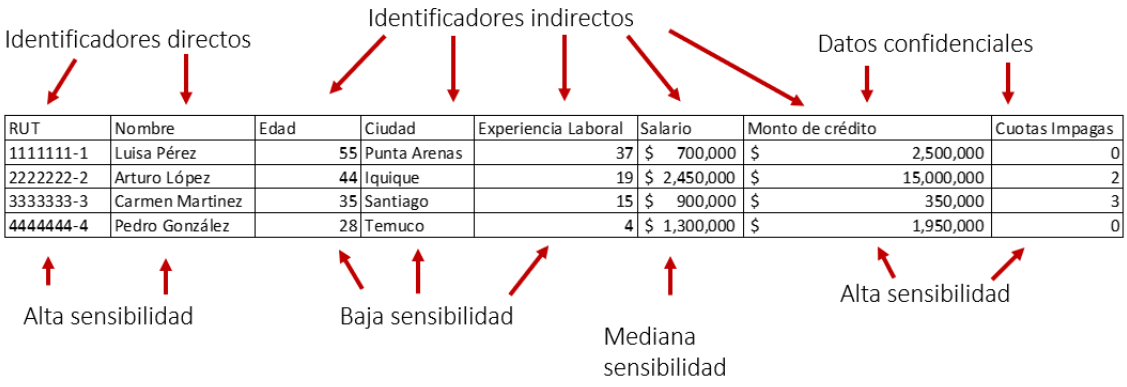
d) Preanonimización y eliminación/reducción de variables

En la preanonimización se determinarán las posibles variables de identificación relevantes para tener en cuenta en el diseño del proceso. Particularmente, en esta etapa se diseña el proyecto de anonimización, en el que se deberán identificar con claridad: i) los identificadores directos e indirectos, ii) los datos confidenciales, iii) la clasificación y sensibilidad de las variables por categoría, iv) identificar potenciales fuentes de información externa que puede ser utilizada para reidentificar los datos; entre otros factores sustanciales que determine la entidad. En la Figura 2 se encuentra una base de datos ficticia que posee distintas variables, tanto personales (RUT, Nombre, Edad, Ciudad, Experiencia laboral) como económicas/financieras del titular (Salario, Monto de crédito total dentro del sistema, Cuotas impagas). Las variables del ejemplo están clasificadas según sus atributos en base a si son indicadores directos/indirectos, confidencialidad y nivel de sensibilidad.

El responsable de la preanonimización, además, deberá identificar y proponer las técnicas adecuadas de anonimización según los atributos del conjunto de datos y el riesgo de reidentificación asociado, para permitir la máxima ruptura de los eslabones en la relación de identificación-información.

Esta etapa debe limitar al mínimo necesario, la cantidad de variables que permitan la identificación de las personas y así reducir el riesgo de reidentificación. Para esto es importante i) Identificar y eliminar tanto las variables directas como indirectas que no vayan a ser necesarias en el tratamiento de los datos anonimizados ii) limitar los plazos de conservación y el uso de la información con exclusivo fin para los que estos fueron solicitados iii) generar un control segregado de usuarios con acceso a los datos personales y usuarios con acceso a los datos anonimizados.

Figura 2. Ejemplo en la determinación de variables de identificación relevantes para el proceso de preanonimización.



RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

e) Selección de las técnicas de anonimización

Las técnicas de anonimización más apropiadas dependerán de la finalidad que se quiere alcanzar con el uso de la información, del tipo de datos y los atributos de la información que se requiere anonimizar. Por lo tanto, se recomienda que la determinación de las técnicas de

anonimización a aplicar sea estudiada y revisada con pruebas que verifiquen su efectividad, desarrollando un proceso interno que puede ser auditado por empresas externas certificadas en la materia. Un ejemplo de ello es la ISO 27001 sobre seguridad de la información.

Existen limitaciones inherentes a las técnicas de anonimización que se deben ponderar y evaluar antes de seleccionar una técnica u otra, atendiendo a los fines previstos para la anonimización. Actualmente, las técnicas de anonimización más utilizadas corresponden a: 1) Generalización de datos, 2) Seudoanonimización, 3) Datos sintéticos, 4) Intercambio/mezcla de datos, 5) Perturbación y 6) Reducción o anulación de data. Sin embargo, debe tenerse en consideración, que el rápido avance tecnológico podría dejar obsoletas las técnicas de anonimización más comunes y reemplazarlas por otras técnicas más seguras. De esta forma, la selección de las técnicas de anonimización debe someterse un análisis continuo de riesgos que ayuden a optar o decidir por las técnicas más apropiadas en cada momento.

f) Anonimización

En la fase de anonimización se realiza la disociación definitiva e idealmente irreversible de los datos personales. No se deberá utilizar un proceso de anonimización de uso general, dado que, para mayor seguridad, este proceso deberá realizarse tantas veces como sea necesario según la finalidad de la información anonimizada, el destinatario y el motivo de tratamiento de estos datos. En esta misma línea, se recomienda que cada conjunto de datos anonimizados generado en cada proceso posea claves distintas, independientemente si el conjunto de datos nominados ya fue anonimizado previamente.

La anonimización es muchas veces un proceso iterativo en el cual, permanentemente se debe ir evaluando la efectividad del proceso, midiendo riesgos y eventualmente, ajustar parámetros o las técnicas de anonimización utilizadas. En este punto, el encargado principal de las iteraciones dentro del proceso de anonimización corresponde al líder o coordinador.

Es altamente recomendable la utilización de algoritmos de sello de tiempo con el fin de garantizar la fecha y hora en la que la anonimización fue realizada, o incluso algoritmos de firma electrónica que permiten garantizar la identidad de quien ha realizado la respectiva anonimización. El objetivo final de la anonimización es proveer los datos desagregados para ser utilizados, sin generar conflictos con los titulares de los datos.

Pasos de la fase de anonimización

i. Determinar los recursos y equipo técnico necesarios para proceder a la anonimización de los datos.

ii. Validar la técnica de anonimización seleccionadas por expertos en etapas anteriores.

iii. Seguir el plan elaborado durante la etapa de preanonimización, donde los encargados deberán aplicar las técnicas seleccionadas, los algoritmos necesarios, realizando pruebas de calidad.

iv. Recodificar o reducir variables para los datos sensibles residuales tras el proceso de anonimización.

v. De ser necesario y pertinente, aplicar técnicas de perturbación de los datos.

vi. Entregar los resultados al responsable del proceso para su aprobación.

vii. Realizar revisiones periódicas del proceso, auditando el uso posterior de los datos mediante métricas o escalas que proporcionen una interpretación objetiva de los resultados.

g) Control

Esta etapa implica la realización de controles periódicos por parte de los servicios tecnológicos de la entidad y con ello, dar soporte al proceso de anonimización, evitando posibles riesgos de reidentificación. Dependiendo del resultado de esta etapa, es posible volver a replantearse distintos escenarios para las etapas previas, para luego volver a realizar nuevos procesos junto a los respectivos controles.

La descripción, características y detalles del proceso de anonimización se deben plasmar en un documento que permita facilitar su revisión, mantenimiento, auditabilidad y replicabilidad. Esta documentación debe estar segura, ya que, contiene información relevante para la reidentificación de la información. Además, debe incorporar la descripción de incidentes en caso en que se hayan presentados y las alternativas de solución implementadas para mitigar el impacto o daño causado.

Estas etapas forman un marco integral que garantiza un proceso estructurado y seguro de anonimización de datos dentro de una institución. Es importante adaptar y personalizar este esquema según las necesidades y contextos específicos de cada organización.

IV. Técnicas de anonimización

Las técnicas de anonimización dependen tanto de los atributos y características de la información, como del tratamiento que se le pretende dar a los datos anonimizados. La elección de la técnica debe ser cuidadosamente estudiada por el equipo de trabajo, tras una evaluación y aprobación previa por parte del equipo evaluador de riesgos, y debe incluir pruebas continuas para verificar su efectividad.

A su vez, el coordinador del proceso de anonimización junto con el equipo evaluador de riesgos deben considerar las limitaciones inherentes a la técnica escogida, junto con los objetivos específicos de la anonimización, con el fin de proteger la privacidad de las personas cuando se publique o consulte cualquier tipo de información dentro de este conjunto de datos.

Las técnicas de anonimización de datos se pueden agrupar según diferentes características del proceso (por ejemplo, técnicas con o sin perturbación) o por enfoques (aleatorización, generalización, seudoanonimización). En la mayoría de los casos, el proceso de anonimización es iterativo y puede ser necesario utilizar varios métodos de anonimización de manera conjunta, dependiendo de las variables y los objetivos específicos.

Las seis técnicas más utilizadas para la anonimización de datos personales son:

1) Generalización de datos:

Técnica que logra la anonimización de un conjunto de datos disminuyendo el nivel de detalle de los valores de los atributos de la información, creando una categorización amplia para brindar una vista generalizada del contenido de los datos. Por lo general, esta técnica es útil cuando se introduce suficiente ambigüedad para lograr los objetivos de privacidad, al tiempo que garantiza que los datos sigan siendo lo suficientemente útiles para su propósito. Aunque la generalización es una herramienta efectiva para descartar la singularización, no puede garantizar la completa anonimización de los datos para todos los casos.

Esta técnica puede realizarse mediante el método de agregación, es decir, generando intervalos de valores para los atributos continuos o creando nuevas supra categorías que contengan a otras más específicas en caso de los atributos categóricos. Esta técnica es ventajosa de utilizar cuando existe un número considerable de individuos que comparten un mismo valor de atributo, dificultando su identificación.

Algunos ejemplos de la generalización mediante la agregación consisten en:

- Agrupar edades específicas en rangos de edad o categorías laborales relacionadas bajo un término general adecuado.
- Omitir la variable de ciudad o comuna, por ejemplo, publicando la variable región que es más agregada.
- El redondeo numérico.

Como se ejemplifica en la Figura 3a, el proceso de generalización básico agrega las observaciones entre aquellos individuos que tienen más de 40 años y aquellos que tienen menos de 40. En este caso, se pierde la información de ciudad, mientras que las variables experiencia laboral, salario, monto total de crédito y cuotas impagas se presentan como el promedio de todos los individuos que pertenecen a cada categoría. El umbral escogido (40 años) es relevante, dado que si se utiliza un corte de 30 años, por ejemplo, aumenta el riesgo de reidentificación es la separación. Como se observa en la Figura 3b, al realizar esta agregación, solamente “Pedro González” sería parte del segundo grupo, por tanto, con alguna base de información anexa sobre ingresos, podría reidentificarse a este individuo. En este sentido, para que la agregación mediante esta técnica tenga mayores índices de seguridad, se debe revisar que ninguna categoría represente a uno o muy pocos individuos identificables.

Finalmente, la Figura 3c muestra un ejemplo de agregación incompleta, en el que se agrupan únicamente la edad y el salario. La edad se clasifica en dos categorías: 40 años o más, y menores de 40 años. Para cada grupo, se presenta el promedio del salario, mientras que los valores originales de las otras variables, que no requieren agregación, se mantienen. Esto resulta en una menor pérdida de información. Aunque esta mayor desagregación puede ser más útil para los fines del estudio, también conlleva un mayor riesgo de reidentificación, especialmente si se utiliza una base de datos adicional que contenga montos de créditos.

Figura 3a. Ejemplo de la técnica de generalización de datos: agregación.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

Rango Edad	Promedio Exp. Labora	Promedio Salaio	Promedio Monto Crédito	Promedio Cuotas Impagas
+40	28.0	$ 1,575,000	$ 8,750,000	1.0
-40	9.5	$ 1,100,000	$ 1,150,000	1.5

Figura 3.b Ejemplo de la técnica de generalización de datos: agregación con mayor riesgo de reidentificación.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

Rango Edad	Promedio Exp. Laboral	Promedio Salaio	Promedio Monto Crédito	Promedio Cuotas Impagas
+30	23.7	$ 1,350,000	$ 5,950,000	1.7
-30	4	$ 1,300,000	$ 1,950,000	0

Figura 3.c Ejemplo de la técnica de generalización de datos: agregación incompleta.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

Obs	Edad	Ciudad	Experiencia Laboral	Promedio Salario	Monto de crédito	Cuotas Impagas
1	+40	Punta Arenas	22	$ 1,575,000	$ 2,500,000	0
2	+40	Iquique	15	$ 1,575,000	$ 15,000,000	2
3	+30	Santiago	10	$ 1,100,000	$ 350,000	3
4	+30	Temuco	6	$ 1,100,000	$ 1,950,000	0

2) Seudoanonimización de datos:

La seudoanonimización de datos se entiende como el proceso de enmascarar información directamente identificable reemplazándola con un identificador artificial, denominado "seudónimo". Normalmente la información que se intercambia son atributos únicos, por lo que, esta técnica no garantiza que el conjunto de datos no pueda someterse a ingeniería inversa si se introduce otro conjunto de datos externo. Las técnicas más comunes que permiten seudoanonimizar corresponden a:

a) **Cifrado con clave secreta**: en este tipo de encriptación, cada individuo posee una clave única la cual se asigna como identificador. Sin embargo, la persona que posee la clave puede revertir fácilmente el proceso mediante la desencriptación.

b) **Algoritmo Hash:** consiste en la utilización de un algoritmo matemático que transforma la información que identifica a la persona en entradas que pueden ser numéricas, de valor o una serie de caracteres de una longitud fija. El algoritmo de Hash permite que, partiendo de un mismo dato se genere siempre la misma "huella digital", pero a la inversa, nunca se pueda obtener el dato original. Este método garantiza la confidencialidad al tratarse de una operación matemática de un solo sentido.

Una alternativa más segura a las anteriores corresponde al algoritmo de Hash con borrado de clave inmediato, donde terminado el proceso de anonimización se procede a borrar la tabla de correspondencia. Con la aplicación de esta técnica es posible reducir la probabilidad de vinculación entre los datos personales de un conjunto de datos y los datos personales del titular contenidos en otro conjunto de datos en el que se ha usado un seudónimo/algoritmo diferente. Sin embargo, para que este método tenga éxito deberá existir un procedimiento que permita la eliminación segura de las claves y la posibilidad de acreditar que el procedimiento se ha cumplido para garantizar la irreversibilidad del proceso.

Los errores y riesgos más frecuentes provenientes del proceso de seudoanonimización consisten en usar la misma clave en varios conjuntos de datos para diferentes conjuntos de usuarios y/o conservar la clave secreta y no guardarla de manera segura.

La Figura 5 presenta ejemplos básicos de la seudoanonimización, tanto usando el método de cifrado con clave numérica (izquierda) como el de Algoritmo de Hash (derecha).

Figura 5. Ejemplo de la técnica de Seudoanonimización.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

Cifrado

ID	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
34	35	Santiago	15	$ 900,000	$ 350,000	3
58	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
50	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
23	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

Algoritmo de Hash

ID	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
0b2d6e7462c57ec22b1289b838e2ca92e20ca3b7	35	Santiago	15	$ 900,000	$ 350,000	3
10144bbfc3d2c141d81a858b1f096cc69379c1fc	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
161f9b9fe444b24f1890f4776e5c16777ecb1ffb	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
5d15dc711083c1815dc5c14be4e6b29c591087cb	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

3) Datos sintéticos:

Los datos sintéticos son datos completamente nuevos generados de forma artificial para sustituir los datos históricos confidenciales, de modo que todos los atributos de datos se modifican de manera muy significativa y todos los registros creados no coinciden con el registro de ningún individuo de los datos originales. De esta manera, la generación de bases de datos sintéticas permite generar archivos de uso público a partir de datos confidenciales/protegidos. A menudo se usa para entornos de prueba y para validar o entrenar modelos matemáticos o de aprendizaje automático, además, de ser útil para modelos de micro simulación o en conjuntos de datos de ejecución remota.

El conjunto de datos sintéticos debe ser realista, es decir, estadísticamente equivalente a la población real de interés. Para ello, la distribución de la población sintética por región y estrato debe ser casi idéntica a la distribución de la población verdadera y las distribuciones marginales y las interacciones entre variables y su correlación deben representarse con precisión. Esta técnica puede utilizar desviaciones estándar, medianas, regresión lineal u otras herramientas estadísticas en la generación de los datos sintéticos.

En la Figura 6, se puede observar que mediante los datos originales se crean 4 observaciones sintéticas que no coincide con ninguna observación real, sin embargo, presentan los mismos promedios para las variables numéricas.

Figura 6. Ejemplo básico para la generación de base de datos utilizando la técnica de Datos sintéticos manteniendo el promedio de la distribución.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0
Promedio		40.5		19	$ 1,337,500	$ 4,950,000	1.3

ID	Edad	Experiencia Laboral	Salario	Monto de crél	Cuotas Impag
1	45	19	$ 960,000	$ 3,000,000	1
2	49	31	$ 2,180,000	$13,500,000	1
3	30	13	$ 850,000	$ 550,000	3
4	38	13	$ 1,360,000	$ 2,750,000	0
Promedio	40.5	19	$ 1,337,500	$ 4,950,000	1.3

4) Intercambio/permutación o mezcla de datos:

El intercambio o permutación de datos implica reposicionar o cambiar los valores de diferentes atributos entre registros, de manera que la información se mantenga útil a nivel agregado, pero sin permitir la identificación de individuos. Esta técnica es pertinente en el caso en que se quiera conservar la distribución exacta de los datos.

El principal riesgo de esta técnica consiste en aplicar la permutación sobre atributos fuertemente correlacionados, ya que, esto contribuye a aumentar la probabilidad de reidentificación. Un ejemplo de lo anterior seria intercambiar información entre individuos sobre los valores del año de nacimiento, años de experiencia e ingresos recibidos manteniendo la relación entre los atributos para cada individuo, tal como se observa en la Figura 7.

Un error potencial al aplicar la permutación de datos es no considerar que intercambiar atributos podría no mitigar el riesgo de reidentificación y que su implementación podría no ofrecer beneficios significativos a la protección de datos personales. En el ejemplo se aprecia que puede ser fácil reidentificar a un sujeto porque las tres variables tienen estrecha correlación entre sí, por lo que, la permutación por sí sola, no permitiría garantizar un adecuado nivel de anonimización y sería necesario reforzar el proceso aplicando otras técnicas de anonimización.

Por otra parte, a pesar de que los datos principales pueden estar anonimizados mediante intercambio o permutación, es posible que se puedan combinar con otros conjuntos de datos disponibles o atributos auxiliares para reidentificar a los individuos. Por ejemplo, datos como la edad, el género y la ubicación geográfica podrían ser suficientes para identificar a personas dentro de un conjunto de datos anonimizado.

Figura 7. Ejemplo básico para la generación de base de datos utilizando la técnica de Intercambio/permutación.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	44	Iquique	19	$ 2.450.000	$ 15.000.000	2
2222222-2	Arturo López	28	Temuco	4	$ 1.300.000	$ 1.950.000	0
3333333-3	Carmen Martinez	55	Punta Arenas	37	$ 700.000	$ 2.500.000	0
4444444-4	Pedro González	35	Santiago	15	$ 900.000	$ 350.000	3

5) Privacidad diferencial o perturbación de datos

La perturbación de datos produce ambigüedad al aleatorizar elementos de los datos y así evitar que las cifras resultantes faciliten información sobre individuos específicos. Por ejemplo, se puede modificar ligeramente el conjunto de datos original agregando ruido aleatorio, transformando las propiedades del conjunto de datos, siendo menos precisos, pero conservando su distribución general.

Cuando esta técnica se realiza de manera deliberada y controlada, los registros individuales podrían ser menos sensibles mientras tienen efectos predecibles y corregibles en el análisis agregado. Sin embargo, aplicar mal esta técnica puede implicar que el ruido agregado esté fuera de escala o que entre los atributos de un conjunto de datos no se presente una lógica congruente.

Mediante la adición de ruido aleatorio, el resultado del análisis se convierte en una aproximación y no en el resultado exacto que se habría obtenido si se hubiera realizado sobre el conjunto de original datos. Sin embargo, esta técnica aplicada de manera correcta no permite que se pueda identificar a

un individuo, ni reparar los datos o detectar cómo se han modificado. Para maximizar su eficacia, se recomienda combinar esta modalidad con otras técnicas de anonimización, como los cuasi identificadores o la eliminación de atributos obvios. En este contexto, la Figura 8 muestra la aplicación de la técnica de perturbación de variables numéricas, junto al algoritmo de Hash aplicado a las variables de identificación directa.

Figura 8 – Ejemplo de perturbación junto a método de Seudoanonimización utilizando algoritmo de Hash.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

ID	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
0b2d6e7462c57ec22b1289b838e2ca92e20ca3b7	57	Punta Arenas	38	$ 722,422	$ 2,580,079	0
10144bbfc3d2c141d81a858b1f096cc69379c1fc	46	Iquique	20	$2,528,477	$ 15,480,471	2
161f9b9fe444b24f1890f4776e5c16777ecb1ffb	34	Santiago	15	$ 871,172	$ 338,789	3
5d15dc711083c1815dc5c14ba4e6b29c591087cb	26	Temuco	4	$1,258,359	$ 1,887,539	0

6) Reducción o anulación de data

Esta técnica consiste en la reducción o eliminación de datos especialmente sensibles que puedan ser identificadores directos, disminuyendo así el nivel de detalle de los datos originales para evitar la presencia de datos únicos. Debido a la pérdida de información que conlleva, su uso generalmente se considera un caso atípico. Se utiliza particularmente cuando es imposible anonimizar a ciertos sujetos, por lo que se deben especificar explícitamente los atributos eliminados y el motivo por el cual se excluyen del resultado final de la anonimización. Por ejemplo, en el caso de tener un individuo con un nivel extremo de renta en comparación al resto de la muestra.

Ejemplos de esta técnica son datos confidenciales como el nombre, la dirección o la edad del cliente que simplemente se eliminan, convirtiendo esta información en valores nulos. La Figura 9 muestra un ejemplo de anulación de data que elimina todas las variables con identificación directa y algunas de identificación indirecta, dejando esta información como "*missing values*".

Figura 9 – Ejemplo de la técnica de anulación de data.

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
1111111-1	Luisa Pérez	55	Punta Arenas	37	$ 700,000	$ 2,500,000	0
2222222-2	Arturo López	44	Iquique	19	$ 2,450,000	$ 15,000,000	2
3333333-3	Carmen Martinez	35	Santiago	15	$ 900,000	$ 350,000	3
4444444-4	Pedro González	28	Temuco	4	$ 1,300,000	$ 1,950,000	0

RUT	Nombre	Edad	Ciudad	Experiencia Laboral	Salario	Monto de crédito	Cuotas Impagas
.	.	.	.	37	$ 700,000	$ 2,500,000	0
.	.	.	.	19	$ 2,450,000	$ 15,000,000	2
.	.	.	.	15	$ 900,000	$ 350,000	3
.	.	.	.	4	$ 1,300,000	$ 1,950,000	0

Capas de anonimización.

Junto a las distintas técnicas de anonimización descritas previamente los procesos de anonimización pueden ser catalogados como monocapa o multicapa. Un proceso de anonimización monocapa es cuando la anonimización de las variables se realiza una única vez y se da por finalizado el proceso. El proceso de denomina multicapa cuando se realizan dos o más procesos de anonimización, simultáneos o secuenciales, al mismo conjunto de datos.

En este contexto, el responsable del proceso ha anonimizado todos los datos personales que puedan servir para reidentificar a los individuos y entrega el conjunto de información innominado a su destinatario en base a lo determinado por la entidad. Quien realice el tratamiento de datos, a fin de evitar minimizar cualquier riesgo remanente de reidentificación, puede decidir realizar un segundo proceso de anonimización de los datos ya anonimizados. De esta forma, el destinatario y/o usuario final del conjunto de información evita que, en caso de fragilidad de los procesos de anonimización previos, la identidad de las personas pudiera verse afectada.

V. Conclusiones

Los procesos de anonimización de datos personales siguen siendo la herramienta más utilizada a nivel internacional para minimizar la probabilidad del mal uso de la información personal, permitiendo así el tratamiento de datos conforme a las leyes y normativas de protección de la vida privada de los titulares de la información. La anonimización tienen un impacto directo en los recursos de una organización (económico, tecnológico, humano, etc.), por lo cual deberán ser adecuados a los objetivos y requerimientos decididos por la entidad.

Sin embargo, estas técnicas presentan limitaciones inherentes debido al rápido avance tecnológico. En el proceso de anonimización, existe un "*trade off*" en lo que respecta a la capacidad tecnológica y procedimiento estadístico de anonimizar y la posibilidad de reidentificar a las personas cuyos datos han sido anonimizados. Es decir, la misma capacidad tecnológica para anonimizar datos personales puede ser utilizada para la reidentificación de los individuos. Este riesgo debe ser considerado como una contingencia latente durante el ciclo de vida de la información.

Por tanto, es necesario reforzar el proceso de anonimización con medidas de evaluación continua de los procesos, la seguridad de la información, los avances tecnológicos, las sanciones y multas ante el uso indebido de datos y, en definitiva, cualquier medida que sirva tanto para atenuar los riesgos de reidentificación de las personas como para paliar las consecuencias de que éstos se materialicen.

En el contexto bancario chileno, una de las finalidades de los procesos de anonimización está relacionado con el uso de modelos internos por parte de la industria financiera. La anonimización permite realizar estimaciones considerando un ciclo económico completo, superando las restricciones legales actuales de 5 años para datos nominados personales relativos a obligaciones de carácter financiero, como lo exige actualmente la Ley sobre Protección de la Vida Privada.

Además, los procesos de anonimización mencionados en esta nota técnica también pueden resultar de gran utilidad en el contexto de la nueva Ley 21.680 sobre Registro de Deuda Consolidada (REDEC), así como en cualquier otra iniciativa legislativa futura relacionada con el tratamiento de datos personales.

VI. Referencias

- Benschop, T., Machingauta, C., & Welch, M. (2019). Statistical disclosure control: A practice guide.
- Brasher, E. A. (2018). Addressing the failure of anonymization: guidance from the European union's general data protection regulation. Colum. Bus. L. Rev., 209.
- Data Protection Commission. (2019). Guidance on Anonymisation and Pseudonymisation. Retrieved November, 7(2019), 2019-06.
- Directiva (UE) 2019/1024 del Parlamento europeo y del consejo de 20 de junio de 2019
- Garrido, J., Bergthaler, M. W., DeLong, M. C. M., Johnson, J., Rasekh, A., Rosha, A., & Stetsenko, N. (2019). The use of data in assessing and designing insolvency systems. International Monetary Fund.
- Instituto Nacional de Estadísticas (2021). “Guía para el control de divulgación estadística en microdatos”. Departamento de Metodologías e Innovación Estadística.
- Lasko, T. A., & Vinterbo, S. A. (2009). Spectral anonymization of data. IEEE transactions on knowledge and data engineering, 22(3), 437-446.
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Boletín Oficial del Estado (BOE) español.
- Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA l. Rev., 57, 1701.
- Reglamento (UE) 2016/679 del Parlamento europeo y del consejo de 27 de abril de 2016.
- Rand, J., & Tarp, F. (2002). Business cycles in developing countries: are they different? World development, 30(12), 2071-2088.