EL MERCURIO Legal

Versión para imprimir El Mercurio.com

La imposibilidad de la anonimización en la nueva ley de protección de datos

"...En la futura litigación en esta materia (...) bastará que un demandante cuente con herramientas estadísticas adecuadas para demostrar que la anonimización de la demandada es reversible. En ese momento, cambia la naturaleza del tipo infraccional y nos encontramos frente a una infracción grave, con sanciones que pueden alcanzar las 20.000 UTM. La pregunta relevante no es cuántos *checklist* se cumplieron, sino si la base de datos resistiría un ataque de reidentificación en un juicio..."

Lunes, 5 de enero de 2026 a las 17:30

Matías Aránguiz y Sebastián Dueñas

Mientras más uno se junta con asesores o empresas intentando adecuarse a la nueva ley de protección de datos (NLDP), más se repite la idea de que si los datos están "anonimizados", la NLDP deja de ser un problema. Esa convicción es falsa por dos razones, una jurídica y otra técnica. La primera es fácil de entender, la segunda se funda en que hace más de 20 años la literatura estadística demuestra que la anonimización es, en la práctica, extraordinariamente frágil.

La letra K) del artículo 3 de la NLDP dice que un dato anonimizado deja de ser personal, pero para que ello ocurra, se debe *hacer por procedimiento irreversible en virtud del cual un dato personal no puede vincularse o asociarse a una persona*. La norma fija una obligación de resultados, no de medios razonables, como en otras jurisdicciones. La única excepción se vislumbra en la anonimización de datos de investigación, donde efectivamente se vislumbra una obligación de medios.

El problema es que esa exigencia de irreversibilidad no se satisface con esfuerzo o estándares de la industria. El problema radica en que si existe una posibilidad objetiva de reconstruir la identidad de un titular, fracasa la anonimización y el dato sigue siendo personal, con todas las consecuencias regulatorias y sancionatorias que ello implica.

Desde el punto de vista estadístico, los datos no identifican personas de forma aislada, sino en combinación: edad, comuna, profesión y fecha de una transacción bastan para reconstruir una identidad concreta con alta probabilidad. Estudios empíricos ampliamente citados muestran que con apenas tres o cuatro atributos "inofensivos" es posible reidentificar a más del 80 % de los individuos en una base de datos supuestamente anónima.

Esto es posible mediante técnicas bien conocidas. Los *linkage attacks* cruzan bases de datos propias con fuentes públicas para encontrar coincidencias únicas. Los *inference attacks* no requieren identificar directamente a una persona, sino que permiten inferir atributos sensibles a partir de patrones agregados. A ello se suman técnicas como *membership inference attacks* o *attribute inference attacks*, hoy ampliamente documentadas.

En 2006 (¡hace 20 años!) Netflix lanzó una competencia en la que investigadores lograron reidentificar usuarios "anónimos" cruzando sus historiales de visualización con reseñas públicas en IMDb. Años antes, Latanya Sweeney demostró que el entonces gobernador de Massachusetts, William Weld, podía ser identificado usando solo código postal, fecha de nacimiento y género. Ninguno de esos datos parecía, por sí solo, particularmente sensible.

La norma chilena no distingue entre los distintos métodos de anonimización, lo que inevitablemente generará confusión en su aplicación. No es lo mismo anonimizar que seudonimizar; no es lo mismo eliminar identificadores directos que garantizar k-anonimidad, cifrar, enmascarar u ofuscar datos. Tampoco es equivalente cumplir la ISO 27559 o la NIST 800-122. Con todo, ni la Agencia ni los tribunales están constreñidos por esa imprecisión legislativa y pueden fundar su interpretación en la evidencia empírica disponible.

En la futura litigación en materia de protección de datos bastará que un demandante cuente con herramientas estadísticas adecuadas para demostrar que la anonimización de la demandada es reversible. En ese momento, cambia la naturaleza del tipo infraccional y nos encontramos frente a una infracción grave, con sanciones que pueden alcanzar las 20.000 UTM. La pregunta relevante no es cuántos *checklist* se cumplieron, sino si la base de datos resistiría un ataque de reidentificación en un juicio.

** Matías Aránguiz Villagrán y Sebastián Dueñas Müller son, respectivamente, director y subdirector del Programa Derecho, Ciencia y Tecnología de la Universidad Católica, además de socios de Inteligencia Digital.*