Desafíos de la Dark Data

Claudia Cardoso[1]

Catherine Collins[2]

La *Dark Data* corresponde a datos olvidados o desconocidos y sin explotar en una organización, que surgen como resultado de las interacciones digitales diarias de los usuarios en línea con innumerables dispositivos y sistemas. Gartner define *Dark Data* como activos de información que las organizaciones recopilan, procesan y almacenan durante sus actividades comerciales habituales, pero que generalmente se utilizan con fines de un solo uso y no se reutilizan para otros fines (por ejemplo, planillas de Excel obsoletas, presentaciones, correos electrónicos, información de excolaboradores o ex cliente, borradores de documentos, datos de sensores de Internet de las cosas que no tienen ningún propósito, etc). Estos datos se generan cada vez en volúmenes más altos y pueden estar en cualquier lugar de la organización.

Según un informe de Quest Software en colaboración con Enterprise Strategy Group, más de la mitad de los datos en las empresas son datos oscuros[3]. Si bien es cierto, la mayoría de las empresas tienen conciencia de que generan este tipo de data, éstas suelen desconocer los riesgos o problemas que puede acarrear. Algunos de los cuales son:

1. Riesgos en protección de datos personales

Según lo establecido en la actual Ley 19.628, el responsable del tratamiento debe cumplir con determinados principios y obligaciones en relación con la recolección y tratamiento de los datos personales, pudiendo destacar que estos deben recolectarse con el consentimiento del titular, o concurriendo otra base de legalidad y para una finalidad determinada, salvo que provengan de una fuente de acceso público. En este sentido podría ocurrir que respecto de estos datos personales oscuros no exista consentimiento u otra base de legalidad para almacenarlos; o, se trate de datos

[1] Gerente de Cumplimiento de Privacidad y Protección de Datos de Falabella.
[2] Abogado Senior Gerencia Cumplimiento Protección de Datos Falabella
[3] 2022 State of Data Governance and Empowerment Report. Disponible en: https://www.erwin.com/analyst-report/2022-state-of-data-governance-and-empowerment-report/ [fecha de visita: 28 de marzo de 2024]

inexactos o caducos y, por tanto, deban rectificarse, anonimizarse o eliminarse. Esto último fue precisamente lo que se le ordenó hacer a Google a propósito de la sentencia que dio a conocer este mes por una demanda colectiva en que se acusaba a esta compañía de recopilar datos de navegación de los usuarios que, en modo incógnito, navegaban en el buscador de Chrome.

Asimismo, el responsable del tratamiento debe cumplir con el principio de proporcionalidad el que consiste en tratar sólo aquellos datos personales que sean estrictamente necesarios, adecuados y pertinentes para alcanzar los fines de tratamiento informados a los titulares al momento de la recolección de los datos, situación que no ocurre con los datos oscuros debido a que por naturaleza son datos desconocidos y por ende no tienen una finalidad clara.

Cabe señalar que el proyecto de ley que busca modificar la ley 19.628, sanciona las situaciones descritas anteriormente con una multa de hasta 10.000 UTM, que puede llegar a 20.000 UTM si concurre dolo o alguna circunstancia agravante.

2. Riesgos de seguridad de la información

Respecto de toda información que sea considerada confidencial, no solo aquella que constituya datos personales, existe la necesidad de resguardar su confidencialidad, integridad y disponibilidad, previniendo accesos o pérdidas no autorizadas, necesidad que puede derivar de una obligación legal o su divulgación o mal uso podría comprometer activos de la compañía o su reputación.

En este sentido, el riesgo se genera porque los datos oscuros suelen no estar sujetos a medidas de seguridad concretas debido a que las empresas no saben dónde se encuentran almacenados, o incluso, que esos datos existen, facilitando de esta forma los ciberataques. Un ejemplo es lo que ocurrió en 2022 con Twitter Bugs. Esta empresa admitió que en algunas cuentas que permanecían conectadas en varios dispositivos móviles, luego de un restablecimiento voluntario de cuentas, las contraseñas se almacenan sin cifrar en logs internos. Esta situación permitió a un “bot” recolectar un número indeterminado de contraseñas y de esa forma robar datos de usuarios de estas cuentas.

En Chile, además de la ley 19.628, existen otros cuerpos legales que exigen resguardar la confidencialidad de la información, entre ellos, la Ley Marco de Ciberseguridad, que establece el deber general de ciberseguridad, por el cual se exige “el establecimiento de medidas para prevenir,

reportar y resolver incidentes de ciberseguridad". Este tipo de obligaciones conlleva la necesidad de determinar y clasificar la información en poder de una empresa, situación que no ocurre con los datos oscuros.

3.- Riesgos al medio ambiente

Los datos oscuros tienen un impacto ambiental significativo ya que, aunque la mayoría de los datos forman parte de la "nube" los servidores donde se almacenan estos requieren grandes cantidades de electricidad para funcionar. Además, el consumo de energía también requiere una cantidad significativa de agua para mantener el nivel de enfriamiento que estos servidores necesitan. Se calcula que los datos oscuros representan más de la mitad de los datos almacenados en el mundo. Este dato no es menor si pensamos que las empresas se preocupan de reducir su huella de carbono sin tener presente sus emisiones producto de sus datos oscuros.

4.- Costos de almacenamiento y pérdida de oportunidad

El almacenamiento de datos oscuros conlleva costos, que si bien hoy no representan montos significativos para las organizaciones en términos de infraestructura de almacenamiento y consumo de recursos de TI, se espera que aumenten considerablemente con el avance tecnológico, por ejemplo, debido al uso de IA y desarrollos de modelos de machine learning por las propias empresas.

Además, los datos oscuros pueden contener información valiosa para las empresas que se está subutilizado debido a que éstas no tienen la capacidad tecnológica o humana para analizar y comprender estos datos.

En conclusión, la dark data genera costos asociados al almacenamiento de información, la pérdida de oportunidad de monetizar esta data, daño al medio ambiente y riesgos relacionados con la seguridad y cumplimiento normativo. Para abordar estos desafíos, las empresas deben implementar estrategias efectivas de gestión de datos que les permitan identificar, clasificar y aprovechar al máximo su data, eliminando aquella información inútil. Esto implica inversiones en tecnologías, capacitaciones en análisis de datos y cultura organizacional orientada al uso y protección de los datos.