Diputació de València | Protecció de Dades i Seguretat de la Informació

Quaderns DivalData

Quaderns dirigits a delegats,
responsables i especialistes en protecció
de dades personals

Quadern N.º 38 | Agost 2023

ERRORS COMUNS EN MATÈRIA D'ANONIMITZACIÓ

ÍNDICE

ERRORS COMUNS EN MATÈRIA D'ANONIMITZACIÓ

Us convidem a traslladar-nos aquelles temàtiques que resulten del vostre interés per als pròxims quaderns. Aquestes peticions hauran de dirigir-se a:

Diputació de València

Departament de Protecció de Dades i Seguretat de la Informació

Pl. de Manises, 4 46003 València

email: dpdsi@dival.es

SUSCRIPCIONS

Si desitges subscriure's al nostre Quadern accedeix al següent enllaç.

“Les dades anònimes constitueixen: «aquella informació que no fa referència a persones naturals identificades o identificables o a dades personals que s'anonimitzen de tal forma que deixen de ser identificables»”.

ÍNTRODUCCIÓ

La **informació anònima** és un conjunt de dades que no guarda relació amb una persona física identificada o identificable (Considerant 26 del Reglament (UE) 2016/679 General de Protecció de Dades, d'ara en avant, RGPD), en tant que la **informació seudonimitzada** és un conjunt de dades que no pot atribuir-se a un interessat sense utilitzar informació addicional, requereix que aquesta informació addicional figure per separat i, a més, estiga subjecta a mesures tècniques i organitzatives destinades a garantir que les dades personals no s'atribuïsquen a una persona física identificada o identificable (Article 4.5 del RGPD).

Transformar un conjunt de dades personals en informació anònima o seudonimitzada exigeix realitzar un tractament sobre aquestes dades personals. El tractament d'anonimització genera un únic i nou conjunt de dades, mentre que el tractament de seudonimització genera dos nous conjunts de dades: la informació seudonimitzada i la informació addicional que permet revertir l'anonimització.

L'anonimització, les seues tècniques, processos i resultats, són qüestions complexes que han generat uns certs dubtes en tant la seua aplicació pràctica.

Per això, a través del present document s'analitzen i resolen les principals confusions que ha generat la tècnica de l'anonimització, explicant la realitat i oferint referències per a una lectura detallada.

Aquests malentesos han sigut resolts per l'Agència Espanyola de Protecció de Dades (d'ara en avant, AEPD) en el seu document anomenat: 10 MALENTESOS RELACIONATS AMB L'ANONIMITZACIÓ.

“REALITAT: «La seudonimització no és el mateix que l'anonimització»”

ERROR 1: «La seudonimització és el mateix que l'anonimització»

La realitat davant aquesta qüestió és que la seudonimització no és el mateix que l'anonimització.

El RGPD defineix «*seudonimització*» en el seu article 4.5) com «*el tractament de dades personals de manera que no puguen atribuir-se a un interessat sense utilitzar informació addicional, sempre que aquesta informació addicional figure per separat i estiga subjecta a mesures tècniques i organitzatives destinades a garantir que les dades personals no s'atribuïsquen a una persona física identificada o identificable*».

Això significa que l'ús de «*informació addicional*» pot suposar la identificació dels individus; per aqueix motiu les dades personals seudonimitzats també són dades personals.

Per contra, les dades anònimes, no poden associar-se amb un individu en particular. Una vegada que les dades són realment anònims i els individus deixen de ser identificables, deixen d'estar inclosos en l'àmbit d'aplicació del RGPD.

ERROR 2: «El xifratge és anonimització»

La realitat es troba en què **el xifratge no constitueix una tècnica d'anonimització**, però pot ser una bona eina de seudonimització.

El procés de xifratge utilitza claus secretes per a transformar la informació de tal forma que es reduïsca el risc d'ús indegut i, al mateix temps, es mantinga la confidencialitat durant un període de temps determinat. Atés que la informació original ha de ser accessible, les transformacions aplicades pels algorismes de xifratge estan dissenyades per a ser reversibles, la qual cosa es coneix com desxifrat.

Les claus privades que s'utilitzen per al desxifrat són la «*informació addicional*», esmentada anteriorment (vegeu Error 1), la qual cosa pot fer que les dades siguen llegibles i, en última instància, que la identificació siga possible.

En teoria, podria considerar-se que l'eliminació de la clau de xifratge de les dades xifrades els convertiria en anònims, però no és així. No es pot donar per fet que les dades xifrades no puguen desxifrar-se perquè es diga que la clau de desxifrat s'ha «*esborrat*» o és «*desconeguda*». Hi ha molts factors que afecten la confidencialitat de les dades xifrades, en particular a llarg termini. Entre aquests factors es troben la solidesa de l'algorisme de xifratge i de la clau, les fugides d'informació, els problemes d'implantació, la quantitat de dades xifrades o els avanços tecnològics (per exemple, la computació quàntica).

REALITAT: El xifratge no constitueix una tècnica d'anonimització, però pot ser una bona eina de seudonimització"

ERROR 3: «Les dades sempre poden anonimitzar-se»

La realitat és que **no sempre és possible reduir el risc de reidentificació** per davall d'un llindar definit de manera prèvia i mantindre, al mateix temps, la utilitat d'un conjunt de dades per a un tractament específic.

L'anonimització és un procés que tracta de trobar l'equilibri adequat entre la reducció del risc de reidentificació i el manteniment de la utilitat d'un conjunt de dades per als fins previstos. No obstant això, en funció del context o la naturalesa de les dades, els riscos de reidentificació podrien no mitigar-se prou. Aquesta situació pot donar-se quan el nombre total de possibles individus («*univers de subjectes*») és massa reduït (per exemple, un conjunt de dades anònimes que continga només els 705 membres del Parlament Europeu), quan les categories de dades són tan diferents entre els individus que és possible individualitzar-los (per exemple, l'empremta digital del dispositiu dels sistemes que van accedir a un determinat lloc web) o quan el cas dels conjunts de dades inclou un elevat nombre d'atributs demogràfics o dades de localització.

"REALITAT: No sempre és possible reduir el risc de reidentificació per davall d'un llindar definit de manera prèvia i mantindre, al mateix temps, la utilitat d'un conjunt de dades per a un tractament específic."

ERROR 4: «L'anonimització és permanent»

La realitat és que **existeix un risc que uns certs processos d'anonimització puguen revertir-se en el futur**. Les circumstàncies poden canviar al llarg del temps i els nous avanços tècnics i la disponibilitat d'informació addicional poden posar en perill els processos d'anonimització previs.

Els recursos informàtics i les noves tecnologies (o nous usos de tecnologies ja existents) disponibles per a un atacant que poguera intentar reidentificar un conjunt de dades anònimes van canviant al llarg del temps. Hui dia, la computació en el núvol proporciona una capacitat de computació assequible a nivells i preus que eren impensables fa anys. En el futur, els ordinadors quàntics també podrien alterar el que en l'actualitat es consideren «*mitjans acceptables*».

A més, la divulgació de dades addicionals al llarg dels anys (per exemple, en una filtració de dades personals) pot permetre que les dades que anteriorment eren anònims s'atribuïsquen a persones identificades. La divulgació de registres de moltes dècades d'antiguitat que continguen dades molt sensibles (per exemple, antecedents penals) podria continuar tenint un efecte bastant perjudicial per a un individu o els seus familiars.

"REALITAT: Existeix un risc que uns certs processos d'anonimització puguen revertir-se en el futur. Les circumstàncies poden canviar al llarg del temps i els nous avanços tècnics i la disponibilitat d'informació addicional poden posar en perill els processos d'anonimització previs."

ERROR 5: «L'anonimització sempre redueix la probabilitat de reidentificació d'un conjunt de dades a zero»

La realitat es troba en què **el procés d'anonimització i la forma en què s'aplique** tindran una influència directa en la probabilitat de **riscos de reidentificació**.

Un procés d'anonimització sòlid té com a objectiu la reducció del risc de reidentificació per davall d'un determinat llindar. Aquest llindar dependrà de diversos factors, com els controls de mitigació existents (cap en el context de la divulgació pública), la repercussió en la privacitat dels individus en cas de reidentificació, els motius i la capacitat d'un atacant per a reidentificar les dades.

Encara que una anonimització del 100% és l'objectiu més desitjable des del punt de vista de la protecció de les dades personals, en alguns casos no és possible i ha de contemplar-se un risc residual de reidentificació.

"REALITAT: El procés d'anonimització i la forma en què s'aplique tindran una influència directa en la probabilitat de riscos de reidentificació"

ERROR 6: «L'anonimització és un concepte binari que no pot mesurar-se»

La realitat és que **el grau d'anonimització pot analitzar-se i mesurar-se**.

L'expressió «*dades anònimes*» no pot entendre's com si els conjunts de dades pogueren etiquetar-se com a anònims o no. Existeix una probabilitat que els registres de qualsevol conjunt de dades es reidentifiquen en funció de la possibilitat d'individualitzar-los. Qualsevol procés sòlid d'anonimització avaluarà el risc de reidentificació, que ha de gestionar-se i controlar-se al llarg del temps.

Excepte en casos específics en els quals les dades estiguen molt generalitzats (per exemple, un conjunt de dades que compte el nombre de visitants d'un lloc web per país en un any), el risc de reidentificació mai pot considerar-se nul.

"REALITAT: El grau d'anonimització pot analitzar-se i mesurar-se."

ERROR 7: «La anonimización puede automatizarse totalmente»

La realitat es troba en què és possible utilitzar eines automàtiques durant el procés d'anonimització, però, donada la importància del context en l'avaluació d'aquest procés, **la intervenció de l'expert humà és necessària**.

Al contrari, requereix una anàlisi del conjunt de dades original, els seus fins previstos, les tècniques que han d'aplicar-se i el risc de reidentificació de les dades resultants.

Malgrat que la identificació i eliminació dels identificadors directes (també coneguda com a «*emmascarament*») constitueix una part important del procés d'anonimització, ha d'anar sempre acompanyada d'una anàlisi cautelosa que busque altres fonts d'identificació (indirecta, en general, a través de quasiidentificadors). Mentre que trobar els identificadors directes és una cosa trivial, els identificadors indirectes, en canvi, no sempre són obvis, i el fet de no detectar-los pot donar lloc a la reversió del procés (és a dir, la reidentificació), la qual cosa té conseqüències per a la privacitat dels individus.

L'automatització podria ser clau en alguns passos del procés d'anonimització, com l'eliminació d'identificadors directes o l'aplicació coherent d'un procediment de generalització sobre una variable. Per contra, sembla poc probable que un procés totalment automatitzat puga identificar quasiidentificadors en diferents contextos o decidir com maximitzar la utilitat de les dades aplicant tècniques específiques a variables concretes.

"REALITAT: És possible utilitzar eines automàtiques durant el procés d'anonimització, però, donada la importància del context en l'avaluació d'aquest procés, la intervenció de l'expert humà és necessària."

“REALITAT: Un procés d'anonimització adequat manté la funcionalitat de les dades per a un fi determinat”.

ERROR 8: «La anonimización inutiliza los datos»

La realitat és que un procés d'anonimització adequat **manté la funcionalitat de les dades per a un fi determinat**.

L'objectiu de l'anonimització és evitar que s'identifique als individus d'un conjunt de dades. Les tècniques d'anonimització sempre restringiran les formes en què es pot utilitzar el conjunt de dades resultant. Per exemple, agrupar les dates de naixement en intervals d'un any reduirà el risc de reidentificació i, al mateix temps, la utilitat del conjunt de dades en alguns casos. Això no significa que les dades anònimes siguen inútils, sinó que la seua utilitat dependrà de la finalitat i del risc de reidentificació que s'accepte.

D'altra banda, les dades personals no poden emmagatzemar-se de manera permanent més del que estipule la seua finalitat original, a l'espera d'una oportunitat en la qual puguen resultar útils per a altres fins. La solució per a alguns responsables del tractament podria ser l'anonimització, en la qual les dades personals poden independitzar-se i rebutjar-se del conjunt de dades, mentre que el conjunt de dades restant continua conservant un significat útil.

El principi de «*minimització de les dades*» exigeix que el responsable del tractament determine si és necessari tractar les dades personals per a complir un objectiu concret, o si aqueix objectiu pot aconseguir-se també amb dades anònimes.

En alguns casos, això pot conduir a la conclusió que l'anonimització de les dades no s'ajusta a la finalitat prevista. En aquests casos, el responsable del tractament haurà de decidir entre tractar les dades personals (i utilitzar, per exemple, la seudonimització) i aplicar el RGPD, o no tractar les dades de cap manera.

ERROR 9: «Seguir un procés d'anonimització que uns altres van utilitzar amb èxit farà que la nostra organització obtinga resultats equivalents»

La realitat es troba en què **els processos d'anonimització han d'adaptar-se al cas concret**, això és, a la naturalesa, l'abast, el context i els fins del tractament, així com als riscos de diversa probabilitat i gravetat per als drets i llibertats de les persones físiques.

L'anonimització no pot aplicar-se com si se seguira una recepta, perquè el context (naturalesa, abast, context i fins del tractament de les dades) probablement difereix d'una circumstància a una altra i d'una organització a una altra. Un procés d'anonimització pot tindre un risc de reidentificació per davall d'un determinat llindar quan les dades només es posen a la disposició d'un nombre limitat de destinataris, mentre que el risc de reidentificació no podrà aconseguir aqueix llindar quan les dades es posen a la disposició del públic en general.

Pot haver-hi diferents conjunts de dades disponibles en diferents contextos. Aquests podrien creuar-se amb les dades anònimes, la qual cosa afectaria el risc de reidentificació. Per exemple, a Suècia, la informació relativa a les dades personals dels contribuents està disponibles de manera pública, mentre que a Espanya no ho estan. Per tant, encara que els conjunts de dades que inclouen informació de ciutadans espanyols i suecs s'anonimitzaren seguint el mateix procediment, els riscos de reidentificació podrien ser diferents.

"REALITAT: Els processos d'anonimització han d'adaptar-se a la naturalesa, l'abast, el context i els fins del tractament, així com als riscos de diversa probabilitat i gravetat per als drets i llibertats de les persones físiques".

“REALITAT: Les dades personals tenen un valor en si mateixos, per als propis individus i per a tercers. La reidentificació d'un individu podria tindre una repercussió greu quant als seus drets i llibertats”.

ERROR 10: «No existeix un risc ni gens d'interés en saber a qui s'atribueixen aquestes dades»

La realitat és que les dades personals tenen un **valor** en si mateixos, per als propis individus i per a tercers. **La reidentificació d'un individu podria tindre una repercussió greu quant als seus drets i llibertats.**

Els atacs contra l'anonimització poden materialitzar-se en forma d'intents deliberats de reidentificació, intents involuntaris de reidentificació, bretxes de seguretat o divulgació de dades al públic. La probabilitat que algú intente reidentificar a un individu només es refereix al primer tipus. No es pot descartar la possibilitat que algú reidentifique almenys a una persona en un conjunt de dades, ja siga per curiositat, per casualitat o per un interés real (per exemple, investigació científica, periodisme o activitat delictiva).

Pot ser difícil avaluar amb precisió l'impacte de la reidentificació en la vida privada d'una persona, perquè sempre dependrà del context i de la informació que es correlacione. Per exemple, la reidentificació d'un interessat en el context aparentment inofensiu de les seues preferències cinematogràfiques podria portar a inferir sobre les inclinacions polítiques o l'orientació sexual d'aqueixa persona.

No obstant això, aquestes dades especialment sensibles gaudeixen d'una protecció especial en virtut del RGPD.

MATERIAL COMPLEMENTARI

- 10 Malentesos relacionats amb l'anonimització publicat per l'Agència Espanyola de Protecció de Dades (AEPD) en aquest enllaç.
- Guia bàsica d'anonimització Elaborada per Autoritat Nacional de Protecció de Dades de Singapur (DPC - Personal Data Protection Commission Singapore) i traduïda per l'Agència Espanyola de Protecció de dades (AEPD) en aquest enllaç.
- Orientacions i garanties en els procediments d'anonimització de dades personals de l'Agència Espanyola de Protecció de dades (AEPD) en aquest enllaç.

NOTICIES

Les autoritats en protecció de dades publiquen diverses notícies d'interés:

1. **La APDCAT posa en marxa la xarxa de delegats i delegades de protecció de dades a Catalunya.**

   L'Autoritat Catalana de Protecció de Dades ha presentat ***'DPD en xarxa'***, la comunitat d'aprenentatge i col·laboració de delegats i delegades dades de Catalunya, impulsada per a afavorir la formació, la comunicació, l'intercanvi d'experiència i el treball en grup entre les persones que vetlen perquè es complisca la normativa de protecció de dades en les organitzacions del sector públic català.

   Consulta la publicació en aquest enllaç.

2. **El projecte de nova Llei Basca de Protecció de Dades personals.**

   El Projecte de Llei de Protecció de Dades Personals i de l'Autoritat Basca de Protecció de Dades, en procés de tramitació en el Parlament Basc, ha sigut aprovat amb l'objectiu de regular el control i la supervisió del tractament de les dades personals dels quals siguen responsables els subjectes inclosos en el seu àmbit d'aplicació.

   El text presenta com a principal novetat la regulació de la "*Autoritat Basca de Protecció de Dades*", que substituirà a l'actual "*Agència Basca de Protecció de Dades*".

   Consulta la publicació en aquest enllaç.