

Dictamen en relació amb la consulta formulada per una fundació del sector públic sobre el desenvolupament d'un model matemàtic per a la prevenció d'epidèmies i pandèmies

Es presenta davant l'Autoritat Catalana de Protecció de Dades un escrit de la Fundació (...) en el qual es demana que l'Autoritat emeti un dictamen sobre el desenvolupament d'un model matemàtic per a la prevenció d'epidèmies i pandèmies a partir de l'ús de dades proporcionades per determinades entitats públiques.

En concret, es planteja si els mecanismes implementats per aquestes entitats garanteixen adequadament el procés d'anonimització del conjunt de dades que cadascuna d'elles proporciona a la Fundació per tal de poder desenvolupar l'esmentat model matemàtic.

La consulta s'acompanya del document "PROJECTE BIG DATA PER LA PREVENCIÓ D'EPIDÈMIES I PANDÈMIES".

Analitzada la consulta, vista la normativa vigent aplicable i d'acord amb l'informe de l'Àrea de Tecnologia i Seguretat de la Informació de l'Autoritat i l'informe de l'Assessoria Jurídica emeto el següent dictamen:

I

(...)

II

La Fundació exposa en la seva consulta que, junt amb altres entitats, està duent a terme un "Projecte" que consisteix en el desenvolupament d'un model matemàtic per a la prevenció d'epidèmies i pandèmies a partir de l'ús de dades proporcionades per determinades entitats públiques.

D'acord amb el document "PROJECTE BIG DATA PER LA PREVENCIÓ D'EPIDÈMIES I PANDÈMIES", adjuntat a la consulta, es tracta, en concret, de desenvolupar un model matemàtic que analitzi les possibles correlacions entre la mobilitat de la ciutadania i la ràtio d'infecció, possibilitant la incidència quantitativa de la mobilitat de la ciutadania amb la incidència de les epidèmies i les condicions meteorològiques.

Aquest model matemàtic es posarà a disposició dels equips d'investigació de la comunitat científica i dels gestors en matèria de sanitat i seguretat pública de les Administracions Públiques que participen en el Projecte.

La fundació afirma en el dit document que no es tracta, en cap cas, d'un projecte de traçabilitat que permeti adoptar decisions sobre individus concrets ni el tractament de dades personals que identifiquin a persones físiques, atès que les dades personals hauran estat anonimitzades i agregades en origen per part de les entitats participants en el Projecte.

En aquest sentit, s'assenyala que el model matemàtic es nodrirà de les dades següents:

- Dades del Programa públic d'anàlítica de dades per a la recerca i la innovació en salut ("PADRIS"), gestionat per l'Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS).
- Dades de la mobilitat a Espanya durant el període de pandèmia per la COVID-19, publicades pel *Ministerio de Transportes, Movilidad y Agenda Urbana* (MITMA).
- Dades obertes del *Instituto Nacional de Estadística* (INE).
- Dades mediambientals, tant meteorològiques com les dels contaminants, facilitades per la Universitat de Girona.

La Fundació fa avinent que en el "Projecte" hi participen les entitats següents:

- a) L'Hospital Germans Trias i Pujol i la Fundació Lluita contra la SIDA, que aporten coneixements científics (epidemiologia).
- b) El *Centre of Innovation for Data tech and Artificial Intelligence* (CDIAI), que desenvoluparà el model de pandèmia per a la prevenció futura.
- c) El *Barcelona Supercomputing Center* (BSC-CNS), que desenvoluparà el model de base de dades per analitzar i monitoritzar casos relacionats amb la pandèmia.
- d) La Universitat Rovira i Virgili (URV), que avaluarà el model matemàtic.
- e) L'Institut de Salut Global de Barcelona (ISGlobal), que aporta assessorament científic en matèria de salut.
- f) L'EURECAT, gestora tècnica del Projecte, que aporta recursos i instal·lacions.

També assenyala que el "Projecte" s'emmarca en les previsions de la Recomanació (UE) 2020/518 de la Comissió Europea de 8 d'abril del 2020, relativa a un conjunt d'instruments comuns de la Unió per a la utilització de la tecnologia i les dades a fi de combatre i superar la crisi de la COVID-19, en particular pel que respecte a les aplicacions mòbils i a la utilització de dades de mobilitat anonimitzades.

A tot això, la Fundació sol·licita a aquesta Autoritat la seva valoració sobre l'adequació dels procediments d'anonimització de les dades a què s'ha fet esment en origen, això és per part de les entitats que en són responsables, per tal de garantir que el model matemàtic objecte del "Projecte" es pot desenvolupar sense generar riscos per a la privacitat de les persones físiques.

III

Fer notar que els principis i garanties de la protecció de dades no s'apliquen a la informació anònima, és a dir, a aquella informació que ha perdut tota vinculació directa o indirecta amb la persona física -o que ja no l'ha tingut des de la seva obtenció-, de manera que l'afectat deixa de ser identificable sense esforços desproporcionats.

Així es desprèn clarament del considerant 26 del Reglament (UE) 2016/679, del Parlament i del Consell Europeu, de 27 d'abril de 2016, General de Protecció de Dades (en endavant, RGPD):

"Los principios de la protección de datos deben aplicarse a toda la información relativa a una persona física identificada o identificable. Los datos personales seudonimizados, que cabría atribuir a una persona física mediante la utilización de información adicional, deben considerarse información sobre una persona física identificable. Para determinar si una persona física es identificable, deben tenerse en cuenta todos los medios, como la singularización, que razonablemente pueda utilizar el responsable del tratamiento o cualquier otra persona para identificar directa o indirectamente a la persona física. Para determinar si existe una probabilidad razonable de que se utilicen medios para identificar a una persona física, deben tenerse en cuenta todos los factores objetivos, como los costes y el tiempo necesarios para la identificación, teniendo en cuenta tanto la

tecnología disponible en el momento del tratamiento como los avances tecnológicos. Por lo tanto los principios de protección de datos no deben aplicarse a la información anónima, es decir información que no guarda relación con una persona física identificada o identificable, ni a los datos convertidos en anónimos de forma que el interesado no sea identificable, o deje de serlo. En consecuencia, el presente Reglamento no afecta al tratamiento de dicha información anónima, inclusive con fines estadísticos o de investigación.”

En la consulta es planteja si les tècniques d’anonimització aplicades per les entitats que proporcionaran les dades que han de servir de fonament per al desenvolupament del model matemàtic per la Fundació i la resta d’entitats participants del “Projecte” garanteixen que ens trobem davant un tractament de dades anonimitzades.

Convé aclarir que qualsevol procediment d’anonimització, aplicat a dades personals, ha de tenir per finalitat destruir el vincle o nexa entre la dada personal i la persona física afectada, a qui es refereix aquesta informació. L’objectiu és que la persona afectada no resulti identificable per tercers sense esforços desproporcionats.

Mentre aquest nexa entre la dada i la persona física a la que aquesta es refereix pugui ser reconstruït de forma relativament senzilla –en aquest sentit, cal considerar-ne tots els factors objectius, com els costos i el temps necessaris per a la identificació, tenint en compte tant la tecnologia disponible en el moment del tractament com els avenços tecnològics-, no es podrà considerar que la informació ha estat objecte d’un procediment d’anonimització adequat i seguirà subjecta als principis i obligacions de la normativa de protecció de dades.

Fer avinent que l’examen de la problemàtica a què es refereix la present consulta s’efectua, tot seguit, tenint en consideració la informació descrita en el document “PROJECTE BIG DATA PER LA PREVENCIÓ D’EPIDÈMIES I PANDÈMIES”, a què s’ha fet esment en el fonament anterior, així com la documentació que se cita en aquest mateix document.

Abans però, convé puntualitzar, atesos els termes en què es formula la consulta, que aquest dictamen se centra en examinar la qualitat de l’anonimització de les dades en origen i la metodologia del “Projecte” en relació amb aquestes dades, als efectes d’avaluar els riscos d’una eventual reidentificació posterior de persones físiques.

Tal com ha fet avinent aquesta Autoritat en dictàmens anteriors (a tall d’exemple, CNS 10/2016, CNS 52/2015, CNS 29/2015, CNS 20/2015 o CNS 34/2014, disponibles al web <https://apdcat.gencat.cat/ca/inici>), en l’entorn del *big data* i amb les possibilitats que ofereixen tècniques com la mineria de dades i la intel·ligència artificial -conceptes als quals es refereix el “Projecte” (apartat 4 del document adjuntat)-, l’encreuament d’informació obtinguda d’origens diversos, fins i tot si ha estat anonimitzada, pot acabar fent identificable una persona.

És a dir, en funció del volum de dades que, en el context del present “Projecte”, es posin a disposició dels ens que hi participen, i segons la forma com s’ofereixin, la possibilitat que la combinació d’aquestes informacions obtingudes de diverses fonts pugui acabar fent identificables persones concretes no s’ha de descartar (a partir de trets generals, el nombre d’individus en la intersecció de tots ells va disminuint fins que s’identifiquen persones concretes).

En aquest mateix sentit s’ha pronunciat el Grup de Treball de l’Article 29 (en endavant, GTA29) en el seu Dictamen 5/2014 sobre tècniques d’anonimització. En aquest dictamen, al qual ens remetem, es fa avinent el següent:

“(…) los responsables del tratamiento deben ser conscientes de que un conjunto de datos anonimizado puede entrañar todavía riesgos residuales para los interesados. Efectivamente, por una parte, la anonimización y la reidentificación son campos de

investigación activos en los que se publican con regularidad nuevos descubrimientos y, por otra, incluso los datos anonimizados, como las estadísticas, pueden usarse para enriquecer los perfiles existentes de personas, con la consiguiente creación de nuevos problemas de protección de datos. En suma, la anonimización no debe contemplarse como un procedimiento esporádico, y los responsables del tratamiento de datos han de evaluar regularmente los riesgos existentes.”

Tal com apunta el GT29 el risc de reidentificació és inherent a qualsevol tècnica d'anonimització, per la qual cosa la intimitat i el dret a la protecció de dades del titular podria veure's compromesa, tot i que les dades hagin estat anonimitzades.

Per aquest motiu, és necessari dur a terme sempre una anàlisi de possibles riscos de reidentificació i, a la vista del resultat obtingut, articular les mesures necessàries per atenuar la probabilitat de que es materialitzin, preveient, fins i tot, mesures reactives per atenuar el possible dany que pogués derivar-se vers una persona física si la dita reidentificació tingués lloc. Aquestes mesures o garanties hauran de ser superiors en aquells casos en què es tractin categories especials de dades (com succeeix en el present cas), atès que el risc és major en atenció al major impacte que representaria aquesta reidentificació, de materialitzar-se, sobre els drets i llibertats de les persones afectades.

Aquesta identificació i anàlisi del risc de reidentificació caldria entendre-la en el present cas com una activitat emmarcada dins l'avaluació d'impacte en la protecció de dades (PIA) a què es refereix l'article 35 del RGPD.

L'RGPD requereix fer una avaluació d'impacte sobre la privacitat *“cuando sea probable que un tipo de tratamiento, en particular si utiliza nuevas tecnologías, por su naturaleza, alcance, contexto o fines, entrañe un alto riesgo para los derechos y libertades de las personas físicas”* (article 35.1). I esmenta expressament com un supòsit en què caldrà fer una avaluació d'impacte, el tractament a gran escala de categories especials de dades, com per exemple, dades de salut (art. 35.2.b)) o l'avaluació sistemàtica i exhaustiva que permeti l'elaboració de perfils (35.2.a)).

En relació amb aquesta avaluació d'impacte, l'LOPDGDD enumera, en l'ur article 28.2, alguns supòsits en què s'entén probable l'existència d'un alt risc per als drets i llibertats de les persones, entre els quals *“cuando se produzca un tratamiento masivo que implique a un gran número de afectados o conlleve la recogida de una gran cantidad de datos personales”* (lletra d).

Tot i que, com s'ha dit, la normativa de protecció de dades no resulta d'aplicació al tractament de dades anònimes i per tant a priori la realització d'una PIA no resultaria en aquest cas exigible, atès que es tracta d'un procediment que busca identificar i controlar el riscs per als drets i les llibertats de les persones associats a un tractament de dades (en aquest cas, de dades anonimitzades) i que, com s'ha vist, el risc de reidentificació és inherent a qualsevol tècnica d'anonimització, el fet que el “Projecte” examinat es fonamenti en l'ús combinat de dades anonimitzades posa de manifest la conveniència, si més no, de la realització en part (no necessàriament hauria d'efectuar-se un procés complet) d'una PIA que permeti mesurar, avaluar i gestionar el risc de reidentificació.

A aquests efectes, pot resultar d'interès consultar la “Guia sobre l'avaluació d'impacte relativa a la protecció de dades al RGPD”, disponible al web de l'Autoritat.

En tot cas, per tal de donar resposta a la present consulta, s'analitza tot seguit, en fonaments jurídics separats, cadascun dels conjunts de dades que es posaran a disposició de la Fundació, i de la resta de participants, per al desenvolupament del “Projecte”, als efectes de determinar la qualitat de l'anonimització que s'ofereix en origen i poder determinar, en base a aquesta, si existeix el risc d'acabar reidentificant les persones afectades.

D'aquesta anàlisi se n'ha exceptuat les dades mediambientals procedents de la Universitat de Girona, atès que aquest tipus d'informació no pot qualificar-se com a dades personals, de conformitat amb l'article 4.1) de l'RGPD.

IV

Avaluació del nivell d'anonimització de les dades de salut

Les dades que aportarà el Departament de Salut al "Projecte" provenen del programa PADRIS, gestionat per l'AQuAS.

En concret, es disposaran de les següents dades:

- Àrea Bàsica de Salut (ABS).
- Rang d'edat.
- Data primera PCR positiva.
- Nombre d'hospitalitzacions.
- Mitjana de dies de hospitalització.
- Nombre d'ingressos en UCI.
- Mitjana de dies d'ingrés en UCI.
- Mortalitat hospitalària.
- Mortalitat no hospitalària a 30 i 60 dies.
- Mortalitat no hospitalària a més de 60 dies.

Les dades de salut especificades posen de manifest que la informació facilitada per a la realització del "Projecte" comprèn diferents informacions de les persones que han patit la COVID-19, les quals es faciliten de manera agregada en funció de l'ABS i del rang d'edat.

Cal però prestar especial atenció al camp "Data primera PCR positiva" atès que, tal com es descriu, és un camp que contindria informació relativa a una persona concreta. Per tant, en àrees amb una petita incidència, no es pot descartar que es pugui acabar identificant una persona a partir de l'edat i el fet d'haver estat diagnosticat positiva.

També els camps "nombre d'hospitalitzacions" i "nombre d'ingressos en UCI", que, tal com es descriuen, semblarien fer referència a una persona concreta.

La resta de camps són dades agregades (mitjanes i totals) a nivell d'ABS i rang d'edat. La qualitat de l'anonimització d'aquests camps depèn del número de persones que contribueixen a cadascun dels grups determinats per l'ABS i el rang d'edat.

En aquest sentit, es desconeix el període temporal que abastaria cada rang d'edat, però és important, cara a reduir el risc de reidentificació, que els rangs d'edat, o fins i tot les franges temporals pel que fa a la data de la primera PCR positiva, siguin suficientment amplis, especialment en aquelles franges en que el nombre d'incidències sigui menor.

En general, atesa la grandària de les ABS (entre 2100 i 50000 persones aprox.) es podria pensar que l'agregació és prou gran i que és difícil obtenir informació d'una persona concreta (vegi's en aquest sentit, les dades de població de referència del Registre central de població del CatSalut, disponibles al web <https://catsalut.gencat.cat/ca/inici/>). Ara bé, tenint en compte que l'agregació també es fa per edat i que la incidència de la COVID-19 és molt variable, no es pot descartar que alguna d'aquestes agregacions faci referència a un nombre reduït de persones.

El factor clau per determinar si les agregacions donen una protecció suficient és el camp "Nombre d'hospitalitzacions". Quan el valor d'aquest camp sigui petit, l'efecte de cadascuna de les persones sobre la resta de camps serà molt alt i, per tant, podria ser possible obtenir informació d'una persona concreta.

A tall d'exemple, si el nombre d'hospitalitzacions és 1, la mitjana de dies d'hospitalització es correspon amb el número de dies que porta hospitalitzada aquesta persona. De forma similar, si el nombre d'hospitalitzacions és 2, qualsevol persona que conegués les dades d'un dels hospitalitzats podria esbrinar-ne les de l'altre.

Als efectes d'evitar aquests riscos, caldria optar per no utilitzar dades quan aquestes fan referència a un grup petit de persones.

Fer notar, en aquest punt, que el punt 7 de la secció 5.4 "Garanties de privacitat de les dades personals" de l'Informe PADRIS -document a què remet la consulta- preveu, com un dels mecanismes que utilitzarà l'AQuAS per anonimitzar en origen les dades del PADRIS, que "s'aplicarà un procés d'anonimització específic per cada treball."

Vist això, no es pot descartar que s'hagi efectuat algun tipus d'anonimització específic (més enllà de l'agregació per ABS i rang d'edat) en relació amb les dades facilitades per al desenvolupament del "Projecte". Ara bé, a l'hora de realitzar aquest dictamen no es té coneixement d'aquest fet.

V

Avaluació del nivell d'anonimització de les dades de mobilitat

Les dades de mobilitat que s'utilitzaran en el "Projecte" són dades obertes proporcionades pel MITMA (<https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opencvata-movilidad>), les quals fan referència a la mobilitat a Espanya durant el període de pandèmia per la COVID-19.

D'acord amb el document "*Análisis de la movilidad en España con tecnología Big Data durante el estado de alarma para la gestión de la crisis del COVID-19*", a què remet la consulta, la font principal d'aquestes dades són registres anonimitzats procedents de les xarxes de telefonia mòbil, la qual s'hauria combinat amb altres fonts de dades per generar matrius origen-destí i altres indicadors de mobilitat i presència de població (dades de usos del sòl, del Padró Municipal d'Habitants i informació de la xarxa de transport), tots ells anonimitzats i agregats.

En aquest mateix document es descriuen les dades emprades, la metodologia i els algorismes d'anàlisi de les dades, així com els indicadors generats.

En essència, com fa avinent la consulta, les dades de mobilitat corresponen al nombre de viatges entre diferents àrees de mobilitat que corresponen a municipis, districtes i agrupacions d'aquests (en el cas de Catalunya, s'han considerat un total de 442 àrees). La població dins aquestes àrees és, en general, superior a 5000 habitants i, en cap cas, inferior a 1000 habitants. Es publiquen dades del nombre de viatges i dels quilòmetres totals agregades per:

- Àrea origen.
- Àrea destí.
- Activitat origen (casa, feina, altres).
- Activitat destí (casa, feina, altres).
- Hora (0, 1, ... , 23).
- Distància (rangs: 0,5-2km, 2-5km, 5-10km, 10-50km, 50-100km i >100km).

Segons es descriu, en el cas que hi hagi una parada superior a una hora, el trajecte ja es passa a comptar com un trajecte diferent.

Convé destacar que, atès que les dades estan basades en la informació reportada per una companyia telefònica, per obtenir els indicadors d'interès, cal extrapolar-les al conjunt de la població. Això es fa expandint les dades d'acord al Padró d'habitants de l'INE i la penetració de l'operador telefònic en cada zona. És a dir, les dades en què es basen les agregacions (nombre de viatges i quilòmetres totals) estan calculades, en part, amb dades reals i, en part, en estimacions fetes a partir del cens i de la penetració de l'operador telefònic.

Existeix el risc que, en cas que una persona tingui un patró de moviments molt únic, això permetés seguir el rastre dels seus moviments.

A tall d'exemple, davant d'un supòsit en què es tingui coneixement que una persona ha fet un viatge poc habitual entre dues poblacions catalanes, amb una distància entre elles d'uns 300 km., en un mateix dia, en què la persona que l'ha fet manifesta no haver fet parades, podem acabar verificant si realment s'ha fet o no el viatge sense parades a partir de la informació facilitada (cal tenir en compte que la informació facilitada inclou l'hora de sortida i arribada, i a més permet saber que no s'han realitzat parades superiors a una hora).

En qualsevol cas, fer notar que, observant les dades publicades pel MITMA, pot constatar-se que sempre que es faciliten en obert dades del nombre de viatges entre dues àrees aquest és superior a 1.000, per la qual cosa el risc esmentat quedaria molt mitigat.

Ara bé, en la descripció metodològica de la construcció de les dades de mobilitat contemplada en el document citat, no hi ha referència a cap mecanisme d'anonimització que només permeti donar dades de mobilitat quan el nombre de viatges entre dos àrees és prou gran, per la qual cosa no es pot afirmar que la publicació de les dades es du a terme sempre amb aquesta garantia.

VI

Avaluació de l'anonimització de les dades sociodemogràfiques

El conjunt de dades de l'INE, publicades com a dades obertes, que es farà servir per al "Projecte" són les següents:

- Renda mitjana per persona (en euros). Mitjana dels anys 2015, 2016 i 2017.
Variable observada a nivell de secció censal.
- Taxa d'atur del 2011 (en percentatges).
Variable observada a nivell de secció censal.
- Percentatge de població de 65 anys o més del 2019.
Variable observada a nivell de secció censal.
- Percentatge d'estrangers de països amb índex de desenvolupament humà mitjà i baix segons el Programa de les Nacions Unides per al Desenvolupament (2019).
Variable observada a nivell de secció censal.
- Infrahabitatge. Percentatge d'habitatges amb menys de 45 m² de superfície habitable de 2011.
Variable observada a nivell de secció censal.
- Densitat de població del 2019 (en habitants/km²).
Variable observada a nivell d'ABS.

A excepció de la densitat de població, totes les altres variables consten observades a nivell de secció censal.

En la consulta s'explica com el Grup de Recerca en Estadística, Econometria i Salut ("GRECS") de la Universitat de Girona ha calculat el valor a nivell d'ABS per a la variable corresponent a la densitat de població. En aquest sentit, s'apunta que utilitzant la població de cadascuna de les seccions censals com a ponderacions (la font de la població total de la secció censal i de la població de la secció censal per sexe va ser INE (2020b)), es va calcular la mitjana ponderada dels valors de les seccions censals que componen l'ABS per obtenir el seu valor a nivell d'ABS.

En atenció a la informació disponible, pot dir-se que en aquest cas no hi hauria risc d'obtenir informació personal a partir d'aquestes dades.

VII

A la vista de les consideracions efectuades en els apartats anteriors, cal concloure que:

Els tres conjunts de dades examinats (dades de salut, dades de mobilitat i dades socioeconòmiques) presenten les dades de forma agregada. Tot i que això redueix significativament el risc de revelar informació personal, és a dir, la possibilitat de reidentificar els afectats. Tot i que a priori no es pot descartar que l'anonimització sigui eficaç en funció dels nivells d'agregació que finalment s'apliquin a cadascun dels subconjunts d'informació, llevat del cas de les dades socioeconòmiques, la informació aportada no ofereix garanties suficients que la informació s'ofereixi amb nivells d'agregació suficientment amplis.

En el cas de les dades de salut, no es pot descartar que la mostra d'individus consistís en una única persona (o en un número molt reduït de persones). Si fos així, les dades publicades serien fàcilment associables a una persona concreta. Tot i que, l'informe PADRIS fa referència a la utilització de mètodes d'anonimització específics a cada projecte, atès que no es disposa d'aquesta informació en relació amb el present "Projecte" no és pot descartar la possibilitat d'agregacions sobre grups petits d'individus.

Per bé que alguna de la informació facilitada podria entendre's, a nivell d'estudi, no especialment crítica (mitjana de dies d'hospitalització, mitjana de dies d'UCI, etc.), cal tenir en consideració que es tracta en tot cas de dades relatives a la salut (article 9 RGPD), i que podria permetre conèixer si una determinada persona ha passat la malaltia, o fins i tot reconstruir algun tipus de seqüència pel que fa a la concatenació de contagis en una determinada àrea, per la qual cosa cal extremar les garanties en el seu tractament, fins i tot, quan ha estat anonimitzada.

El risc esmentat tampoc pot descartar-se en el cas de les dades de mobilitat, atès que, si bé s'ha observat en alguns dels fitxers de dades publicats que les dades dels viatges només s'ofereixen en obert quan l'agregació correspon a més de 1000, aquesta observació no queda fixada en la metodologia d'anonimització feta pel MITMA.

Per tot plegat, atesos els riscos detectats de reidentificació per correlació amb altres conjunts de dades, per tal de poder dur a terme el "Projecte" caldria adoptar les mesures oportunes per atenuar les possibles conseqüències que poguessin derivar-se en cas que s'arribés a materialitzar la reidentificació de persones físiques.

En aquest sentit, caldria plantejar-se si el desenvolupament del model matemàtic a què es refereix la consulta requereix disposar inexcusablement de la totalitat de dades examinades (PADRIS, Open data mobilitat i sociodemogràfiques).

La necessària aplicació del principi de minimització recollit a l'article 5.1.c) RGPD (tractar la informació personal mínima imprescindible) és clau quan es tracten dades personals, però també si es fa un tractament de dades anonimitzades.

El tractament de dades que poden resultar excessives o innecessàries pot incrementar considerablement el risc de reidentificació, per la qual cosa, des de la vessant de la protecció de dades, cal limitar sempre les dades tractades a les mínimes indispensables per a assolir la finalitat pretesa, tant des del punt de vista del volum de població (nombre de registres) com de dades analitzades (atributs processats).

També caldria maximitzar el nivell d'agregació i davant mostres d'individus petites o extremadament reduïdes, emascarar les dades relatives a aquests col·lectius o bé optar per eliminar-les.

Al seu torn, caldria oferir garanties addicionals per tal de preservar els drets dels afectats, la qual cosa implicaria una sèrie de compromisos per part de tots els participants en el "Projecte":

- a) Deure de confidencialitat.
- b) Mantenir l'anonimització, és a dir, no fer accions per reidentificar.
- c) Vincular la informació lliurada exclusivament al "Projecte", no destinant-la a altres propòsits.
- d) Comunicar immediatament a les entitats que proporcionen les dades qualsevol sospita de reidentificació (a aquests efectes convindria establir un protocol per poder dur a terme aquesta comunicació d'una forma àgil i segura).
- e) Establir un termini màxim de conservació de les dades i destruir-les un cop ja no siguin necessàries pel "Projecte" (la qual cosa caldria acreditar davant les entitats que proporcionen les dades).

D'acord amb les consideracions fetes fins ara en relació amb la consulta plantejada, es fan les següents,

Conclusions

Per la informació de què es disposa, no es pot descartar que a partir l'encreuament de les dades anonimitzades a què es refereix la consulta per tal de desenvolupar un model matemàtic per a la prevenció d'epidèmies i pandèmies es pogués acabar identificant persones concretes, per la qual cosa cal mesurar, avaluar i gestionar aquest risc de reidentificació adoptant les mesures adequades, a què s'ha fet referència, per reduir la probabilitat de reidentificació.

Barcelona, 23 de març de 2021