

**Dictamen en relación a la consulta formulada por una fundación del sector público sobre el desarrollo de un modelo matemático para la prevención de epidemias y pandemias**

**Se presenta ante la Autoridad Catalana de Protección de Datos un escrito de la Fundación (...) en el que se pide que la Autoridad emita un dictamen sobre el desarrollo de un modelo matemático para la prevención de epidemias y pandemias a partir del uso de datos proporcionados por determinadas entidades públicas.**

**En concreto, se plantea si los mecanismos implementados por estas entidades garantizan adecuadamente el proceso de anonimización del conjunto de datos que cada una de ellas proporciona a la Fundación para poder desarrollar el citado modelo matemático.**

**La consulta se acompaña del documento “PROYECTO BIG FECHA POR LA PREVENCIÓN DE EPIDEMIAS Y PANDEMIAS”.**

**Analizada la consulta, vista la normativa vigente aplicable y de acuerdo con el informe del Área de Tecnología y Seguridad de la Información de la Autoridad y el informe de la Asesoría Jurídica emito el siguiente dictamen:**

(...)

II

**La Fundación expone en su consulta que, junto a otras entidades, está llevando a cabo un “Proyecto” que consiste en el desarrollo de un modelo matemático para la prevención de epidemias y pandemias a partir del uso de datos proporcionados por determinadas entidades públicas.**

**De acuerdo con el documento “PROYECTO BIG FECHA POR LA PREVENCIÓN DE EPIDEMIAS Y PANDEMIAS”, adjuntado a la consulta, se trata, en concreto, de desarrollar un modelo matemático que analice las posibles correlaciones entre la movilidad de la ciudadanía y el ratio de infección, posibilitando la incidencia cuantitativa de la movilidad de la ciudadanía con la incidencia de las epidemias y condiciones meteorológicas.**

**Este modelo matemático se pondrá a disposición de los equipos de investigación de la comunidad científica y de los gestores en materia de sanidad y seguridad pública de las administraciones públicas que participan en el Proyecto.**

**La fundación afirma en dicho documento que no se trata, en ningún caso, de un proyecto de trazabilidad que permita adoptar decisiones sobre individuos concretos ni el tratamiento de datos personales que identifiquen a personas físicas, dado que los datos personales habrán sido anonimizados y agregadas en origen por parte de las entidades participantes en el proyecto.**

**En este sentido, se señala que el modelo matemático se nutrirá de los siguientes datos:**

- Datos del Programa público de analítica de datos para la investigación y la innovación en salud (“PADRIS”), gestionado por la Agencia de Calidad y Evaluación Sanitarias de Cataluña (AQuAS).
  - Datos de la movilidad en España durante el período de pandemia por la COVID-19, publicados por el Ministerio de Transportes, Movilidad y Agenda Urbana (MITMA).
  - Datos abiertos del Instituto Nacional de Estadística (INE).
- Datos medioambientales, tanto meteorológicos como los de los contaminantes, facilitados por la Universidad de Girona.

La Fundación recuerda que en el “Proyecto” participan las siguientes entidades:

- a) El Hospital Germans Trias i Pujol y la Fundación Lucha contra el SIDA, que aportan conocimientos científicos (epidemiología).
- b) El Centro of Innovation for Data tech and Artificial Intelligence (CDIAI), que desarrollará el modelo de pandemia para la prevención futura.
- c) Barcelona Supercomputing Center (BSC-CNS), que desarrollará el modelo de base de datos para analizar y monitorizar casos relacionados con la pandemia.
- d) La Universidad Rovira i Virgili (URV), que evaluará el modelo matemático.
- e) El Instituto de Salud Global de Barcelona (ISGlobal), que aporta asesoramiento científico en materia de salud.
- f) EL EURECAT, gestora técnica del Proyecto, que aporta recursos e instalaciones.

También señala que el “Proyecto” se enmarca en las previsiones de la Recomendación (UE) 2020/518 de la Comisión Europea de 8 de abril de 2020, relativa a un conjunto de instrumentos comunes de la Unión para la utilización de la tecnología y los datos a fin de combatir y superar la crisis de la COVID-19, en particular en lo que respecta a las aplicaciones móviles ya la utilización de datos de movilidad anonimizados.

A todo ello, la Fundación solicita a esta Autoridad su valoración sobre la adecuación de los procedimientos de anonimización de los datos a los que se ha mencionado en origen, esto es por parte de las entidades que son responsables, por para garantizar que el modelo matemático objeto del “Proyecto” puede desarrollarse sin generar riesgos para la privacidad de las personas físicas.

### III

Hacer notar que los principios y garantías de la protección de datos no se aplican a la información anónima, es decir, a aquella información que ha perdido toda vinculación directa o indirecta con la persona física -o que ya no la ha tenido desde de su obtención, de modo que el afectado deja de ser identificable sin esfuerzos desproporcionados.

Así se desprende claramente del considerante 26 del Reglamento (UE) 2016/679, del Parlamento y del Consejo Europeo, de 27 de abril de 2016, General de Protección de Datos (en adelante, 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 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 se plantea si las técnicas de anonimización aplicadas por las entidades que proporcionarán los datos que deben servir de fundamento para el desarrollo del modelo matemático por la Fundación y el resto de entidades participantes del “Proyecto” garantizan que nos encontramos ante un tratamiento de datos anonimizados.

Conviene aclarar que cualquier procedimiento de anonimización, aplicado a datos personales, tendrá por finalidad destruir el vínculo o nexo entre el dato personal y la persona física afectada, a quien se refiere esta información. El objetivo es que la persona afectada no resulte identificable por terceros sin esfuerzos desproporcionados.

Mientras este nexo entre el dato y la persona física a que éste se refiere pueda ser reconstruido de forma relativamente sencilla –en este sentido, hay que considerar todos sus 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-, no se podrá considerar que la información ha sido objeto de un procedimiento de anonimización adecuado y seguirá sujeta a los principios y obligaciones de la normativa de protección de datos.

Recuerda que el examen de la problemática a que se refiere la presente consulta se efectúa, a continuación, teniendo en consideración la información descrita en el documento “PROYECTO BIG FECHA PARA LA PREVENCIÓN DE EPIDEMIAS Y PANDEMIAS”, a las que se ha mencionado en el fundamento anterior, así como la documentación que se cita en este mismo documento.

Antes, conviene puntualizar, dados los términos en los que se formula la consulta, que este dictamen se centra en examinar la calidad de la anonimización de los datos en origen y la metodología del “Proyecto” en relación con estos datos, a efectos de evaluar los riesgos de una eventual reidentificación posterior de personas físicas.

Tal y como ha recuerdo esta Autoridad en dictámenes anteriores (a modo de ejemplo, CNS 10/2016, CNS 52/2015, CNS 29/2015, CNS 20/2015 o CNS 34/2014, disponibles en la web <https://apdcat.gencat.cat/ca/inici>), en el entorno del big data y con las posibilidades que ofrecen técnicas como la minería de datos y la inteligencia artificial -conceptos a los que se refiere el “Proyecto” (apartado 4 del documento adjuntado)-, el cruce de información obtenida de orígenes diversos, incluso si ha sido anonimizada, puede acabar haciendo identificable a una persona.

Es decir, en función del volumen de datos que, en el contexto del presente “Proyecto”, se pongan a disposición de los entes que participan, y según la forma en que se ofrezcan, la posibilidad de que la combinación de estas informaciones obtenidas de varias fuentes pueda acabar haciendo identificables personas concretas no debe descartarse (a partir de rasgos generales, el número de individuos en la intersección de todos ellos va disminuyendo hasta que se identifican personas concretas).

En este mismo sentido se ha pronunciado el Grupo de Trabajo del Artículo 29 (en adelante, GTA29) en su Dictamen 5/2014 sobre técnicas de anonimización. En este dictamen, al que nos remitimos, se recuerda lo siguiente:

“(…) los responsables del tratamiento deben ser conscientes de que un conjunto de datos anonimizado puede entrañar aún 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 deben evaluar regularmente los riesgos existentes.”

Tal y como apunta el GT29, el riesgo de reidentificación es inherente a cualquier técnica de anonimización, por lo que la intimidad y el derecho a la protección de datos del titular podría verse comprometida, aunque los datos hayan sido anonimizados.

Por este motivo, es necesario llevar a cabo siempre un análisis de posibles riesgos de reidentificación y, a la vista del resultado obtenido, articular las medidas necesarias para atenuar la probabilidad de que se materialicen, previendo, incluso, medidas reactivas para atenuar el posible daño que pudiera derivarse hacia una persona física si dicha reidentificación tuviera lugar. Estas medidas o garantías tendrán que ser superiores en aquellos casos en que se traten categorías especiales de datos (como sucede en el presente caso), dado que el riesgo es mayor en atención al mayor impacto que representaría esta reidentificación, de materializarse, sobre los derechos y libertades de las personas afectadas.

Esta identificación y análisis del riesgo de reidentificación debería entenderse en el presente caso como una actividad enmarcada dentro de la evaluación de impacto en la protección de datos (PIA) a que se refiere el artículo 35 del RGPD.

El RGPD requiere realizar una evaluación de impacto sobre la privacidad “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” (artículo 35.1). Y menciona expresamente como un supuesto en el que habrá que realizar una evaluación de impacto, el tratamiento a gran escala de categorías especiales de datos, como por ejemplo, datos de salud (art. 35.2.b)) o la evaluación sistemática y exhaustiva que permita la elaboración de perfil

En relación con esta evaluación de impacto, la LOPDDDD enumera, en su artículo 28.2, algunos supuestos en los que se entiende probable la existencia de un alto riesgo para los derechos y libertades de las personas, entre los cuales “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” (letra

Aunque, como se ha dicho, la normativa de protección de datos no resulta de aplicación al tratamiento de datos anónimos y por tanto a priori la realización de una PIA no resultaría en este caso exigible, dado que se trata de un procedimiento que busca identificar y controlar los riesgos para los derechos y libertades de las personas asociados a un tratamiento de datos (en este caso, de datos anonimizados) y que, como se ha visto, el riesgo de reidentificación es inherente a cualquiera técnica de anonimización, el hecho de que el “Proyecto” examinado se fundamente en el uso combinado de datos anonimizados pone de manifiesto la conveniencia, al menos, de la realización en parte (no necesariamente debería efectuarse un proceso completo) de una PIA que permita medir, evaluar y gestionar el riesgo de reidentificación.

A estos efectos, puede resultar de interés consultar la “Guía sobre la evaluación de impacto relativa a la protección de datos en el RGPD”, disponible en la web de la Autoridad.

En cualquier caso, para dar respuesta a la presente consulta, se analiza a continuación, en fundamentos jurídicos separados, cada uno de los conjuntos de datos que se pondrán a disposición de la Fundación, y del resto de participantes, para el desarrollo del “Proyecto”, a efectos de determinar la calidad de la anonimización que se ofrece en origen y poder determinar, en base a ésta, si existe el riesgo de acabar reidentificando a las personas afectadas.

De este análisis se han exceptuado los datos medioambientales procedentes de la Universidad de Girona, dado que este tipo de información no puede calificarse como datos personales, de conformidad con el artículo 4.1) del RGPD.

#### IV

##### Evaluación del nivel de anonimización de los datos de salud

Los datos que aportará el Departamento de Salud al “Proyecto” provienen del programa PADRIS, gestionado por AQUAS.

En concreto, se dispondrán de los siguientes datos:

- Área Básica de Salud (ABS).
- Rango de edad.
- Fecha primera PCR positiva.
- Número de hospitalizaciones.
- Media de días de hospitalización.
- Número de ingresos en UCI.
- Media de días de ingreso en UCI.
- Mortalidad hospitalaria.
- Mortalidad no hospitalaria a 30 y 60 días.
- Mortalidad no hospitalaria a más de 60 días.

Los datos de salud especificados ponen de manifiesto que la información facilitada para la realización del “Proyecto” comprende diferentes informaciones de las personas que han sufrido la COVID-19, las cuales se facilitan de forma agregada en función del ABS y del rango de edad.

Sin embargo, hay que prestar especial atención al campo “Fecha primera PCR positiva” dado que, tal y como se describe, es un campo que contendría información relativa a una persona concreta. Por tanto, en áreas con una pequeña incidencia, no se puede descartar que se pueda acabar identificando a una persona a partir de la edad y el hecho de haber sido diagnosticado positiva.

También los campos "número de hospitalizaciones" y "número de ingresos en UCI", que, tal y como se describen, parecerían hacer referencia a una persona concreta.

El resto de campos son datos agregados (medias y totales) a nivel de ABS y rango de edad. La calidad de la anonimización de estos campos depende del número de personas que contribuyen a cada uno de los grupos determinados por el ABS y el rango de edad.

En este sentido, se desconoce el período temporal que abarcaría cada rango de edad, pero es importante, cara a reducir el riesgo de reidentificación, que los rangos de edad, o incluso las franjas temporales con respecto a la fecha de la primera PCR positiva, sean suficientemente amplios, especialmente en aquellas franjas en las que el número de incidencias sea menor.

Por lo general, dado el tamaño de las ABS (entre 2100 y 50000 personas aprox.) se podría pensar que la agregación es lo suficientemente grande y que es difícil obtener información de una persona concreta (véase en este sentido, los datos de población de referencia del Registro central de población del CatSalut, disponibles en la web <https://catsalut.gencat.cat/ca/inici/>). Ahora bien, teniendo en cuenta que la agregación también se realiza por edad y que la incidencia de la COVID-19 es muy variable, no puede descartarse que alguna de estas agregaciones haga referencia a un número reducido de

El factor clave para determinar si las agregaciones dan una protección suficiente es el campo "Número de hospitalizaciones". Cuando el valor de este campo sea pequeño, el efecto de cada una de las personas sobre el resto de campos será muy alto y, por tanto, podría ser posible obtener información de una persona concreta.

A modo de ejemplo, si el número de hospitalizaciones es 1, la media de días de hospitalización se corresponde con el número de días que lleva hospitalizada esta persona. De forma similar, si el número de hospitalizaciones es 2, cualquier persona que conociera los datos de uno de los hospitalizados podría averiguar los del otro.

A efectos de evitar estos riesgos, habría que optar por no utilizar datos cuando éstos se refieren a un grupo pequeño de personas.

Hacer notar, en este punto, que el punto 7 de la sección 5.4 "Garantías de privacidad de los datos personales" del Informe PADRIS -documento al que remite la consulta- prevé, como uno de los mecanismos que utilizará AQuAS para anonimizar en origen los datos del PADRIS, que "se aplicará un proceso de anonimización específico para cada trabajo."

Visto esto, no puede descartarse que se haya efectuado algún tipo de anonimización específico (más allá de la agregación por ABS y rango de edad) en relación con los datos facilitados para el desarrollo del "Proyecto". Ahora bien, a la hora de realizar este dictamen no se tiene conocimiento de este hecho.

v

#### Evaluación del nivel de anonimización de los datos de movilidad

Los datos de movilidad que se utilizarán en el "Proyecto" son datos abiertos proporcionados por el MITMA ([https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata movilidad](https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/.opendata.movilidad)), las cuales hacen referencia a la movilidad en España durante el periodo de pandemia por la COVID-19.

De acuerdo con el documento "Análisis de la movilidad en España con tecnología Big Data durante el estado de alarma para la gestión de la crisis del COVID-19", al que remite la consulta, la fuente principal de estos datos son registros anonimizados procedentes de las redes de telefonía móvil, que se habría combinado con otras fuentes de datos para generar matrices origen-destino y otros indicadores de movilidad y presencia de población (datos de usos del suelo, del Padrón Municipal de Habitantes e información de la red de transporte), todos ellos ano

En este mismo documento se describen los datos empleados, la metodología y los algoritmos de análisis de los datos, así como los indicadores generados.

En esencia, como recuerda la consulta, los datos de movilidad corresponden al número de viajes entre diferentes áreas de movilidad que corresponden a municipios, distritos y agrupaciones de éstos (en el caso de Cataluña, se han considerado un total de 442 áreas ). La población en estas áreas es, en general, superior a 5000 habitantes y, en ningún caso, inferior a 1000 habitantes. Se publican datos del número de viajes y de los kilómetros totales agregados p

- Área origen. •
- Área destino. •

Actividad origen (casa, trabajo, otros).

- Actividad destino (casa, trabajo, otros). • Hora (0, 1, ..., 23). • Distancia

(rangos: 0,5-2km, 2-5km, 5-10km, 10-50km, 50-100km y >100km).

Según se describe, en caso de que haya una parada superior a una hora, el trayecto ya se pasa a contar como un trayecto diferente.

Conviene destacar que, dado que los datos están basados en la información reportada por una compañía telefónica, para obtener los indicadores de interés, es necesario extrapolarlos al conjunto de la población. Esto se hace expandiendo los datos de acuerdo al Padrón de habitantes del INE y la penetración del operador telefónico en cada zona. Es decir, los datos en los que se basan las agregaciones (número de viajes y kilómetros totales) están calculados, en parte, con datos reales y, en parte, en estimaciones realizadas a partir del censo y de la penetración del operador.

Existe el riesgo de que, en caso de que una persona tuviera un patrón de movimientos muy único, esto permitiera seguir el rastro de sus movimientos.

A modo de ejemplo, ante un supuesto en el que se tenga conocimiento de que una persona ha realizado un viaje poco habitual entre dos poblaciones catalanas, con una distancia entre ellas de unos 300 km., en un mismo día, en el que la persona que lo ha hecho manifiesto no haber hecho paradas, podemos acabar verificando si realmente se ha realizado o no el viaje sin paradas a partir de la información facilitada (hay que tener en cuenta que la información facilitada incluye la hora de salida y llegada, y además permite saber que no se han realizado paradas superiores a una hora).

En cualquier caso, hacer notar que, observando los datos publicados por el MITMA, puede constatarse que siempre que se faciliten en abierto datos del número de viajes entre dos áreas éste es superior a 1.000, por lo que dicho riesgo quedaría muy mitigado.

Ahora bien, en la descripción metodológica de la construcción de los datos de movilidad contemplada en el citado documento, no hay referencia a ningún mecanismo de anonimización que sólo permita dar datos de movilidad cuando el número de viajes entre dos áreas es suficientemente grande, por lo que no puede afirmarse que la publicación de los datos se lleva a cabo siempre con esta garantía.

## VI

### Evaluación de la anonimización de los datos sociodemográficos

El conjunto de datos del INE, publicados como datos abiertos, que se utilizará para el "Proyecto" son los siguientes:

- Renta media por persona (en euros). Media de los años 2015, 2016 y 2017.  
Variable observada a nivel de sección censal.
- Tasa de paro de 2011 (en porcentajes).  
Variable observada a nivel de sección censal.
- Porcentaje de población de 65 o más años de 2019.  
Variable observada a nivel de sección censal.
- Porcentaje de extranjeros de países con índice de desarrollo humano medio y bajo según el Programa de Naciones Unidas para el Desarrollo (2019).  
Variable observada a nivel de sección censal.
- Infravivienda. Porcentaje de viviendas con menos de 45 m<sup>2</sup> de superficie habitable de 2011.  
Variable observada a nivel de sección censal.
- Densidad de población de 2019 (en habitantes/km<sup>2</sup>).  
Variable observada a nivel de ABS.

A excepción de la densidad de población, todas las demás variables constan observadas a nivel de sección censal.

En la consulta se explica cómo el Grupo de Investigación en Estadística, Econometría y Salud (“GRECS”) de la Universidad de Girona ha calculado el valor a nivel de ABS para la variable correspondiente a la densidad de población. En este sentido, se apunta que utilizando la población de cada una de las secciones censales como ponderaciones (la fuente de la población total de la sección censal y de la población de la sección censal por sexo fue INE (2020b)), se calculó la media ponderada de los valores de las secciones censales que componen el ABS para obtener su valor a nivel de ABS.

En atención a la información disponible, puede decirse que en este caso no habría riesgo de obtener información personal a partir de estos datos.

## VII

A la vista de las consideraciones efectuadas en los apartados anteriores, cabe concluir que:

Los tres conjuntos de datos examinados (datos de salud, datos de movilidad y datos socioeconómicos) presentan los datos de forma agregada. Aunque esto reduce significativamente el riesgo de revelar información personal, es decir, la posibilidad de reidentificar a los afectados. Aunque a priori no puede descartarse que la anonimización sea eficaz en función de los niveles de agregación que finalmente se apliquen a cada uno de los subconjuntos de información, salvo en el caso de los datos socioeconómicos, la información aportada no ofrece garantías suficientes que la información se ofrezca con niveles de agregación suficientemente amplios.

En el caso de los datos de salud, no puede descartarse que la muestra de individuos consistiera en una única persona (o en un número muy reducido de personas). Si así fuera, los datos publicados serían fácilmente asociables a una persona concreta. Aunque el informe PADRIS hace referencia a la utilización de métodos de anonimización específicos en cada proyecto, dado que no se dispone de esta información en relación con el presente “Proyecto” no se puede descartar la posibilidad de agregaciones sobre grupos pequeños de individuos.

Si bien alguna de la información facilitada podría entenderse, a nivel de estudio, no especialmente crítica (media de días de hospitalización, media de días de UCI, etc.), es necesario tener en consideración que se trata en todo caso de datos relativos a la salud (artículo 9 RGPD), y que podría permitir conocer si una determinada persona ha pasado la enfermedad, o incluso reconstruir algún tipo de secuencia en cuanto a la concatenación de contagios en una determinada área, por lo que es necesario extremar las garantías en su tratamiento, incluso, cuando ha sido anonimizada.

El mencionado riesgo tampoco puede descartarse en el caso de los datos de movilidad, dado que, si bien se ha observado en algunos de los ficheros de datos publicados que los datos de los viajes sólo se ofrecen en abierto cuando la agregación corresponde a más de 1000, esta observación no queda fijada en la metodología de anonimización realizada por el MITMA.

Por todo ello, dados los riesgos detectados de reidentificación por correlación con otros conjuntos de datos, para poder llevar a cabo el “Proyecto” habría que adoptar las medidas oportunas para atenuar las posibles consecuencias que pudieran derivarse en caso de que se llegara a materializar la reidentificación de personas físicas.

En este sentido, habría que plantearse si el desarrollo del modelo matemático al que se refiere la consulta requiere disponer inexcusablemente de la totalidad de datos examinados (PADRIS, Open data mobilitat y sociodemográficas).



La necesaria aplicación del principio de minimización recogido en el artículo 5.1.c) RGPD (tratar la información personal mínima imprescindible) es clave cuando se tratan datos personales, pero también si se realiza un tratamiento de datos anonimizados.

El tratamiento de datos que pueden resultar excesivos o innecesarios puede incrementar considerablemente el riesgo de reidentificación, por lo que, desde la vertiente de la protección de datos, es necesario limitar siempre los datos tratados a los mínimos indispensables para alcanzar la finalidad pretendida, tanto desde el punto de vista del volumen de población (número de registros) como de datos analizados (atributos procesados).

También habría que maximizar el nivel de agregación y ante muestras de individuos pequeñas o extremadamente reducidas, enmascarar los datos relativos a estos colectivos o bien optar por eliminarlos.

A su vez, habría que ofrecer garantías adicionales para preservar los derechos de los afectados, lo que implicaría una serie de compromisos por parte de todos los participantes en el "Proyecto":

- a) Deber de confidencialidad.
- b) Mantener la anonimización, es decir, no realizar acciones para reidentificar.
- c) Vincular la información entregada exclusivamente al "Proyecto", no destinándola a otros propósitos.
- d) Comunicar de inmediato a las entidades que proporcionan los datos cualquier sospecha de reidentificación (a estos efectos convendría establecer un protocolo para poder llevar a cabo esta comunicación de una forma ágil y segura).
- e) Establecer un plazo máximo de conservación de los datos y destruirlos una vez ya no sean necesarios para el "Proyecto" (lo cual debería acreditarse ante las entidades que proporcionan los datos).

De acuerdo con las consideraciones hechas hasta ahora en relación con la consulta planteada, se hacen las siguientes,

### Conclusiones

Por la información de que se dispone, no se puede descartar que a partir del cruce de los datos anonimizados a los que se refiere la consulta para desarrollar un modelo matemático para la prevención de epidemias y pandemias se pudiera acabar identificando a personas concretas, por lo que es necesario medir, evaluar y gestionar este riesgo de reidentificación adoptando las medidas adecuadas, a las que se ha hecho referencia, para reducir la probabilidad de reidentificación.

Barcelona, 23 de marzo de 2021