

INVESTIGACIÓN BIOMÉDICA, *BIG DATA* Y PROTECCIÓN DE DATOS

Investigación biomédica, big data y protección de datos

La aplicación de técnicas de *big data* en investigación biomédica puede facilitar grandes avances en las ciencias de la salud. Sin embargo, su uso requiere del diseño y la aplicación de medidas de privacidad desde el diseño que garanticen el cumplimiento de la normativa de protección de datos.

A estos efectos, se analiza la nueva normativa general y la sectorial española en materia de protección de datos para identificar los principales retos y algunas herramientas útiles para acometer esta tarea y permitir el aprovechamiento del potencial de la tecnología y el respeto al derecho fundamental de la protección de datos.

PALABRAS CLAVE

Protección de datos, Big data, Investigación biomédica, Salud, RGPD, LOPDGDD

Biomedical research, big data and data protection

The application of big data techniques in biomedical research can lead to great advances in health sciences. However, its use requires the design and implementation of privacy-by-design measures which guarantee compliance with data protection regulations.

For these purposes, the new general privacy regulations and the Spanish sectorial regulations regarding data protection are analyzed in order to identify the main challenges and some useful tools to undertake this task and exploit the potential of technology respecting data protection, which is a fundamental right.

KEY WORDS

Data Protection, Big data, Biomedical research, Health, GDPR, LOPDGDD

Fecha de recepción: 5-9-2019

Fecha de aceptación: 9-9-2019

1 · INTRODUCCIÓN

Las ciencias de la salud se han basado tradicionalmente en la observación y la experiencia. El conocimiento médico acumulado tras cientos de años de estudio de pacientes y casos particulares ha permitido, lenta y trabajosamente, llegar a formular reglas generales que aplicar en el tratamiento médico.

Sin embargo, el uso de las tecnologías de la información y la digitalización en el entorno sanitario han permitido, en cuestión de pocas décadas, obtener información y compartir conocimiento de pacientes y dolencias en todo el mundo con una agilidad cada vez mayor, así como un seguimiento mucho más preciso y personalizado de su evolución.

Por tanto, la incorporación de técnicas de *big data* a la investigación biomédica puede constituir el siguiente salto cualitativo en la evolución del conocimiento médico. Pese a que no existe una única definición del término *big data*, con carácter general hace referencia al conjunto de tecnologías, algoritmos y sistemas empleados para recolectar datos a un *volumen*, *variedad* y *velocidad* nunca alcanzadas hasta hace pocos años, así como a la extracción de información de *valor* de dichos datos mediante sistemas de analítica avanzada.

En el contexto biomédico, el uso de técnicas de *big data* permitiría extraer e incorporar al ámbito de la investigación un mayor volumen de información,

de naturaleza y fuentes distintas de las tradicionalmente utilizadas (por ejemplo, procedente de *wearables* o de fuentes no estructuradas y, potencialmente, de cualquier lugar del mundo) y a una velocidad cada vez más cercana al tiempo real.

A la luz de lo anterior, parece evidente que el uso de técnicas de *big data* podría contribuir no solo a enormes avances en el ámbito de la investigación biomédica, sino también a mejorar la efectividad y eficiencia de los servicios sanitarios y a racionalizar los recursos dedicados al sector de la salud.

La oportunidad se ha detectado a nivel global, y durante la última década han proliferado los estudios e iniciativas relacionados con el tema. En este sentido, resulta particularmente interesante el Estudio sobre *big data* en sanidad pública, telemedicina y atención médica (*Study on Big Data in Public Health, Telemedicine and Healthcare*) realizado por la Comisión Europea y cuyo informe final fue publicado en diciembre de 2016.

Sin perjuicio de los potenciales beneficios para la ciencia médica (y para la humanidad en su conjunto) de la aplicación de técnicas de *big data* en el ámbito de la investigación biomédica, las entidades que las empleen deben asegurar que su uso respete la normativa aplicable y los derechos fundamentales en juego, entre los que cobra especial relevancia el derecho a la protección de datos personales.

2 · NORMATIVA APLICABLE

El diseño de un proyecto de investigación biomédica requiere, como es natural, de la identificación y análisis previo del marco normativo aplicable al proyecto. En este sentido, la normativa de protección de datos española y europea (en particular, el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos —“RGPD”— y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales —“LOPDGDD”—) resultará aplicable materialmente (y deberá analizarse, además, su ámbito de aplicación territorial) en la medida en que el desarrollo del proyecto requiera del tratamiento de datos personales.

Lo anterior implica que el RGPD y la LOPDGDD únicamente resultarán de aplicación al tratamiento de datos personales en el contexto del proyecto de investigación. Por tanto, en la medida en que la información que se trate (a) no se refiera a una persona física identificada o identificable, o (b) tras la implementación de un proceso de anonimización no pueda vincularse la información con la persona física a la que pertenece, los principios y exigencias de estas normas no resultarán de aplicación.

Sin perjuicio de lo anterior, lo cierto es que, en la práctica, es muy complicado que un proyecto de investigación biomédica no requiera en absoluto del tratamiento y análisis de información que pueda, aun potencialmente, vincularse con el sujeto al que pertenece (teniendo en cuenta los factores objetivos previstos en el considerando 26 del RGPD).

En consecuencia, lo habitual será que al menos parte de la información que se trate en el marco del proyecto incluya datos personales que, como veremos a continuación, podrán ser datos seudonimizados (esto es, sometidos a un proceso como resultado del cual ya no puedan atribuirse a un interesado sin utilizar información adicional), pero que estarán sometidos, en todo caso, a la normativa de protección de datos.

Asimismo, será también habitual que los datos personales tratados en el marco de la investigación biomédica tengan la consideración de categorías especiales de datos por tratarse de información relativa a la salud, lo que conllevará la aplicación de medidas de protección reforzadas que deberán tenerse en cuenta.

Por último, la normativa general en materia de protección de datos deberá aplicarse conjuntamente con la normativa sectorial aplicable (con carácter general, la Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica —“LAP”—, y la Ley 14/2007, de 3 de julio, de investigación biomédica —“LIB”—).

3 · PRINCIPALES RETOS

Sobre la base de las anteriores consideraciones, a continuación se exponen algunos de los principales retos, desde la perspectiva de la protección de datos, que deberían abordarse al diseñar un proyecto de investigación biomédica en el que se empleen técnicas de *big data*.

3.1 · Licitud del tratamiento

El tratamiento de datos personales en el marco de la investigación biomédica suele basarse en el consentimiento expreso del sujeto para una investigación concreta. Sin embargo, puede darse el caso (muy habitualmente en proyectos de *big data*) de que muestras biológicas o información asistencial obtenidas para una finalidad específica en el marco de una investigación concreta pretendan utilizarse en otro momento en un proyecto de investigación diferente.

En estos supuestos, la necesidad de obtener nuevos consentimientos específicos para cada investigación concreta constituye un obstáculo susceptible de entorpecer el aprovechamiento de las oportunidades que ofrece el uso de técnicas de *big data* en investigación biomédica.

Si acudimos a la normativa sectorial, la LIB (modificada por última vez en el año 2011) prevé algunas excepciones al principio general de consentimiento específico para una investigación concreta, si bien se circunscriben a que (a) la obtención del nuevo consentimiento no sea posible o represente un esfuerzo no razonable (previo dictamen favorable del Comité Ético de Investigación, que debería tener en cuenta los requisitos previstos en el artículo 58.2 de la LIB), o a que (b) el consentimiento específico obtenido ya previera el empleo de la muestra para otras líneas de investigación relacionadas con la inicialmente propuesta.

En la misma línea, y en relación con la información asistencial contenida en la historia clínica, el artículo 16.3 de la LAP prevé que el acceso a la historia clínica con fines de investigación se realice de manera que se preserven los datos de identificación personal del paciente separados de los de carácter clínicoasistencial, de manera que, como regla general, quede asegurado el anonimato, salvo que el propio paciente hubiera consentido no separarlos, debiendo respetarse en todo caso la normativa aplicable (en particular, la de protección de datos).

Sin embargo, este artículo ha sido modificado con ocasión de la entrada en vigor de la LOPDGDD, con el objeto de exceptuar de su ámbito de aplicación los supuestos de investigación previstos en el apartado segundo de la disposición adicional decimoséptima de la propia LOPDGDD, a la que nos referiremos a continuación.

En efecto, la normativa sectorial debe interpretarse conjuntamente con los principios y disposiciones del RGPD y la LOPDGDD, en particular en lo relativo a la licitud del tratamiento de datos personales en el marco de la investigación científica (en particular, la biomédica), aspecto que estas normas se ocupan de regular.

No es casual que tanto el RGPD como la LOPDGDD recojan disposiciones concretas en esta materia. Ello es coherente con la pretensión de promocionar un mercado único y cohesionado en materia de investigación en el seno de la Unión Europea, así como de cohonestar el respeto al derecho a la protección de datos y el desarrollo científico a nivel europeo.

Asimismo, la Agencia Española de Protección de Datos (“AEPD”) ha tenido ocasión de interpretar la normativa sectorial española referente a investigación biomédica a la luz del RGPD en su Informe Jurídico 073667/2018. La principal conclusión alcanzada por la AEPD es que los requisitos de especificidad y carácter inequívoco para la prestación del consentimiento en materia de protección de datos no deben ser interpretados en el ámbito de la investigación científica de un modo restrictivo.

Así, sobre la base del artículo 89 del RGPD en relación con los artículos 9.2.a) y j) (este último interpretado conjuntamente con el artículo 58.2 de la LIB) y varios considerandos del RGPD (en particular, el 33), considera que es posible prestar un consentimiento válido en relación con un determinado campo de investigación, *“pudiendo extenderse en el futuro ese consentimiento, sin que ello lo vicie en modo*

alguno, incluso a “finalidades” o áreas de investigación que ni siquiera hubieran podido determinarse en el momento en que se prestó sin que sea necesario recabar un nuevo consentimiento del sujeto fuente, teniendo en cuenta los beneficios para los individuos y la sociedad en su conjunto que pueden derivarse de tal investigación no prevista”.

Concluye la AEPD que un consentimiento prestado en relación con una rama amplia de investigación, como por ejemplo la investigación oncológica (o incluso ámbitos más extensos), podría considerarse suficientemente inequívoco y específico.

Esta interpretación es coherente con lo dispuesto en el apartado segundo de la ya mencionada disposición adicional decimoséptima de la LOPDGDD, que incluye las siguientes disposiciones que pueden contribuir a facilitar el tratamiento de grandes conjuntos de datos personales en investigación biomédica:

- Los fines de investigación en salud (y, en particular, la biomédica) para los que el interesado preste el consentimiento podrán abarcar categorías relacionadas con áreas generales vinculadas a una especialidad médica o investigadora.
- La reutilización de datos personales con fines de investigación en materia de salud y biomédica se considera lícita y compatible cuando, habiéndose obtenido el consentimiento para una finalidad concreta, se utilicen los datos para finalidades o áreas de investigación relacionadas con el área en la que se integrase científicamente el estudio inicial (previo informe favorable del comité de ética y publicación de la información prevista en el artículo 13 del RGPD, como se indica en el apartado 3.2 siguiente).

Las previsiones anteriores, unidas al reconocimiento expreso por parte del RGPD (en su considerando 157) de que la combinación de información procedente de registros puede permitir a los investigadores obtener resultados más sólidos (al basarse en una población mayor) y nuevos conocimientos de gran valor sobre condiciones médicas extendidas, pueden contribuir a facilitar la implementación de técnicas de *big data* en la investigación biomédica.

Sin perjuicio de lo anterior, un análisis exhaustivo del proyecto y de su base de licitud con carácter previo a su implementación será fundamental para garantizar la licitud del tratamiento de datos personales durante la fase de implementación del proyecto de investigación.

3.2 · Transparencia e información al interesado

El principio de transparencia es uno de los pilares fundamentales del sistema de protección de datos y uno de los grandes retos que plantea el uso de técnicas de *big data* en investigación biomédica, precisamente por el gran volumen de información tratada, la variedad en su naturaleza y origen y la velocidad a la que tienen lugar los tratamientos.

Asimismo, en el sector de la salud las obligaciones de transparencia previstas en la normativa general de protección de datos se solapan con las reguladoras de los derechos y la autonomía del paciente, así como con las previstas en la normativa específica de investigación biomédica (en particular, en el artículo 15 de la LIB). Por ello, las organizaciones deberán esforzarse por diseñar e implementar mecanismos que permitan cumplir con las obligaciones legales en materia de información y transparencia, lo que se traduce en el establecimiento de medidas adecuadas para garantizar el derecho de información del interesado en relación con el tratamiento, uso y destino de sus datos personales.

En este sentido, la disposición adicional decimoséptima de la LOPDGDD pone también el foco en la necesidad de que el tratamiento de datos personales en el marco del proyecto de investigación se lleve a cabo de forma transparente frente al interesado.

Así, en los supuestos de reutilización de datos personales para finalidades o áreas de investigación relacionadas con el área en la que se integrase científicamente el estudio inicial para el que el interesado prestó su consentimiento (y en relación con el cual recibió la información pertinente), la norma permite que la información establecida por el artículo 13 del RGPD sea publicada en un lugar fácilmente accesible de la página web corporativa del centro donde se realice la investigación o estudio clínico, y, en su caso, en la del promotor, y que dicha circunstancia sea notificada por medios electrónicos a los afectados.

Esta clase de medidas pueden contribuir a facilitar el cumplimiento de sus obligaciones en materia de transparencia a las organizaciones, si bien estas deben emplear esfuerzos suficientes para garantizar que los titulares de los datos personales que tratan puedan recibir la información pertinente de manera completa y transparente.

3.3 · Conservación de la información

El uso de técnicas de *big data* conlleva la gestión de enormes conjuntos de datos de distintas fuentes y, habitualmente, la conservación y reutilización de la información obtenida. Esta tendencia puede ser contraria a algunos de los principios y obligaciones específicas en materia de protección de datos y, en particular, el principio de minimización.

La preocupación que esta materia genera en el legislador comunitario queda patente en la redacción del artículo 89 del RGPD, que hace hincapié en la necesidad de que el tratamiento con fines de investigación científica quede sujeto a garantías adecuadas que hagan que se disponga de medidas técnicas y organizativas, en particular para garantizar el respeto del principio de minimización de los datos personales.

Sin perjuicio de lo anterior, si bien la norma general es que los datos personales se conserven durante el tiempo imprescindible para cumplir con la finalidad que motivó su recogida, lo cierto es que el artículo 5.1.e) del RGPD prevé la posibilidad de conservar los datos durante periodos más largos siempre que se traten con fines de investigación científica, debiendo aplicar medidas técnicas y organizativas adecuadas (por ejemplo, la seudonimización, como se prevé en el considerando 156).

Los principios anteriores deberán ser analizados en el momento de diseñar el proyecto y durante su implementación, a fin de garantizar que la información personal, seudonimizada o no, no se conserve (al menos, sin anonimizar) durante más tiempo del estrictamente necesario en relación con los fines de investigación científica.

3.4 · Derechos del interesado

Otro de los grandes retos asociados a los proyectos de *big data* (tanto en el ámbito biomédico como en otros sectores) es la gestión de los derechos de los interesados en materia de protección de datos, no solo por la dificultad práctica que puede suponer en este contexto, sino porque, en determinadas ocasiones, su ejercicio puede llegar a comprometer la viabilidad del proyecto.

Consciente de esta realidad, la disposición adicional decimoséptima de la LOPDGDD prevé la posibilidad de excepcionar los derechos de acceso, rectificación, limitación del tratamiento y oposición, si bien limitada al cumplimiento de una serie de

requisitos, entre los que se incluye el que la investigación tenga por objeto un interés público esencial relacionado con la seguridad del Estado, la defensa, la seguridad pública u otros objetivos importantes de interés público general (en este último caso, la excepción deberá constar en una norma con rango de Ley).

En los demás supuestos, será necesario diseñar e implementar mecanismos que permitan el respeto de los derechos de los interesados recogidos en los artículos 15 a 22 del RGPD, de manera que no se comprometa la viabilidad del proyecto.

4 · PRIVACIDAD DESDE EL DISEÑO

Una vez expuestos algunos de los retos a los que se enfrenta la implementación de técnicas de *big data* en investigación biomédica, a continuación se enuncian algunas herramientas que pueden resultar de utilidad para acometer el proyecto garantizando el respeto a la normativa de protección de datos y desde una perspectiva de privacidad desde el diseño.

4.1 · Seudonimización

Como se ha adelantado, aunque el análisis de información anónima excluiría la aplicación de la normativa de protección de datos, es en la práctica muy complicado acometer un proyecto de investigación sin necesidad de tratar información que no pueda ser vinculada a una persona física identificada o identificable.

Sin embargo, sí es posible seudonimizar los datos personales de manera que no puedan atribuirse al interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física.

En este sentido, la disposición adicional decimoséptima de la LOPDGDD considera lícito el uso de datos personales seudonimizados con fines de investigación en salud y, en particular, biomédica. A continuación enumera una serie de requisitos aplicables al uso de datos personales seudonimizados con fines de investigación en salud pública y biomédica (no queda clara la intención del legislador al cualificar la investigación en salud como “pública” en este punto; en particular porque los requisitos que se enuncian a continuación debe-

rían aplicarse, razonablemente, a cualquier proceso de seudonimización en el ámbito sanitario y biomédico):

1.º Una separación técnica y funcional entre el equipo investigador y quienes realicen la seudonimización y conserven la información que posibilite la reidentificación.

2.º Que los datos seudonimizados únicamente sean accesibles al equipo de investigación cuando (i) exista un compromiso expreso de confidencialidad y de no realizar ninguna actividad de reidentificación; y (ii) se adopten medidas de seguridad específicas para evitar la reidentificación y el acceso de terceros no autorizados.

En definitiva, la seudonimización de datos no solo constituye una medida adicional de respeto al derecho a la protección de datos de los interesados, sino que también puede ser de utilidad para el desarrollo del proyecto, al facilitar la operativa para el tratamiento de datos personales.

4.2 · Evaluación de impacto en la protección de datos

Como norma general, el tratamiento de datos que se lleve a cabo en el marco de una investigación biomédica en la que se empleen técnicas de *big data* requerirá de la realización de una evaluación de impacto de conformidad con el artículo 35 del RGPD, al tratarse de un tipo de tratamiento que utiliza nuevas tecnologías y que entraña el tratamiento a gran escala de categorías especiales de datos (criterios que, además, han sido específicamente incluidos en el listado de tratamientos en los que es obligatorio realizar una evaluación de impacto, publicado por la AEPD en mayo de 2019).

Adicionalmente, la disposición adicional decimoséptima de la LOPDGDD prevé que en aquellos casos en que se lleve a cabo un tratamiento con fines de investigación en salud pública y, en particular, biomédica deberá procederse a la realización de una evaluación de impacto que deberá incluir de modo específico los riesgos de reidentificación vinculados a la anonimización o seudonimización de los datos (como se ha indicado con anterioridad, no queda claro qué aporta la cualificación como “pública” de la investigación en salud en esta disposición; ¿implica que los tratamientos con fines de investigación en salud en general no requieren necesariamente de la realización de una evaluación de impacto y del cumplimiento de los demás requi-

sitos previstos, como someter la investigación a normas de calidad sobre buena práctica clínica?).

Independientemente de que la realización de una evaluación de impacto se considere obligatoria o no, lo cierto es que su realización resulta recomendable y de gran utilidad para garantizar que el proyecto cumpla con los principios y disposiciones de la normativa de protección de datos.

4.3 · Delegado de protección de datos

Al igual que en el caso de la evaluación de impacto, la involucración en el proyecto de un experto en privacidad como el delegado de protección de datos resulta muy conveniente para garantizar el respeto de la normativa de protección de datos durante la gestión del proyecto, en particular si en este se emplean técnicas de *big data* y tiene lugar en el ámbito de la salud.

Adicionalmente, la LOPDGDD prevé que el uso de datos personales seudonimizados con fines de investigación en salud pública y, en particular, biomédica deberá ser sometido al informe previo del comité de ética de la investigación previsto en la normativa sectorial o, en su defecto, el informe previo del delegado de protección de datos o de un experto con los conocimientos previos en el artículo 37.5 del RGPD, lo que refuerza la conveniencia de contar con asesoramiento especializado en proyectos de estas características.

5 · CONCLUSIONES

En la medida en que los proyectos de investigación biomédica (incluidos aquellos en los que se empleen técnicas de *big data*) conlleven el trata-

miento de datos de carácter personal, deberá cumplirse con los principios y obligaciones específicos previstos en la normativa general de protección de datos y en la normativa sectorial que resulte de aplicación.

Como ha quedado patente en la exposición, el RGPD y la LOPDGDD contienen principios y disposiciones específicas (tanto en materia de transparencia e información al interesado como de base de licitud o de gestión de derechos de los interesados) que, interpretados conjuntamente con la normativa sectorial española, pueden permitir el cumplimiento de la normativa de protección de datos sin obstaculizar el progreso de la investigación biomédica.

Asimismo, herramientas como la seudonimización y la realización de evaluaciones de impacto en la privacidad o la figura del delegado de protección de datos pueden ser de gran ayuda tanto en la fase de preparación como durante la implementación y revisión del proyecto.

En definitiva, es necesario que las organizaciones que gestionen proyectos de investigación biomédica (en particular, aquellos que incorporen tecnologías disruptivas como técnicas de *big data* o inteligencia artificial) adopten un enfoque de *accountability* y de privacidad desde el diseño que permita afrontar los retos que se presentan para el respeto a la privacidad.

Solo de esta manera podrá aprovecharse el potencial de las nuevas tecnologías para el desarrollo de la investigación biomédica y la mejora de la salud y la vida de las personas sin poner en riesgo el derecho (fundamental, no lo olvidemos) a la protección de datos de las personas implicadas en el proceso.

ANE IRUNE LAMELA DOMÍNGUEZ (*)

(*) Abogada del Área de Mercantil de Uría Menéndez (Bilbao).