



EDITORIAL

Big data y seguridad de la información

Joan Gené Badia*, Pedro Gallo de Puelles e Itziar de Lecuona

Universidad de Barcelona, Barcelona, España

Cuando todavía no se ha conseguido abandonar el paternalismo médico para lograr que los pacientes adopten sus decisiones, Yoav Noah Harari en su «bestseller» *Homo Deus*¹ afirma que en un futuro cercano preferiremos que las máquinas tomen las decisiones importantes por nosotros. Seguramente ambos procesos son compatibles ya que el autor considera que voluntariamente optaremos por la opción electrónica. Cuando elegimos, lo hacemos procesando los datos que disponemos a través de un algoritmo. Sabemos que las decisiones son más acertadas cuanto mayor y mejor es la información utilizada. Es posible que los ordenadores escojan mejor que nosotros porque emplean algoritmos más sofisticados y procesan mayores volúmenes de datos. Actualmente no solo cuentan con la mejor evidencia científica disponible, sino que también poseen un profundo conocimiento de nosotros mismos, incluidas nuestras preferencias, valores y formas de pensar. Por el momento no hemos conseguido la infalibilidad de estos procesos automáticos, todavía pueden estar sesgados y llevarnos a conclusiones erróneas.

A lo largo del día dejamos un rastro electrónico de nuestras actividades. Utilizando el ordenador, llevando teléfono móvil o pagando con la tarjeta de crédito generamos una información que nos describe con bastante detalle. Un estudio de las Universidades de Cambridge y Stanford muestra que un algoritmo electrónico automático sin participación humana que analizaba solo los *likes* emitidos en Facebook® permitía predecir la personalidad de un individuo mejor que sus amigos, familiares o pareja². No sorprende, por tanto, que estas grandes bases de datos de imágenes, opiniones, compras, viajes o caminatas estén consideradas el nuevo recurso natural. Empresas como Google®, Twitter® o Facebook® ofrecen servicios gratuitos para conseguir que les

facilemos voluntariamente, una valiosa información que luego comercializan con fines que desconocemos y para los que no fuimos consultados.

Estos grandes volúmenes de datos, habitualmente complejos, que no pueden procesarse por métodos tradicionales son los denominados *big data*. La tecnología *big data* es capaz de manejar a gran velocidad grandes volúmenes de información valiosa, compleja, interconectada y de tipología muy variada, que puede utilizarse para tomar decisiones³⁻⁶. Sin embargo, la garantía de la veracidad de los datos y la ausencia de sesgos solo queda asegurada por la capacidad humana de gestionar el proceso⁷.

Es por ello, que la ingente cantidad de información generada y almacenada en el entorno sanitario brinda una excelente oportunidad para la tecnología *big data*. Al utilizar un identificador único podemos relacionar la información clínica con datos generados por la actividad humana (consumos, ocio, actividad física o intercambio de información). El conjunto describe con increíble precisión nuestra salud y los factores que la condicionan. El *big data* abre nuevos campos de investigación y aporta un instrumento inestimable para adoptar con más precisión decisiones clínicas individuales.

La Unión Europea promovió que para 2015 todos los estados miembros dispusieran de una legislación que facilitara la reutilización de la información pública para fines distintos de los que la originaron (*Public Sector Information Directive* 2003/98/EC). Siguiendo estas directrices, España publicó los Reales Decretos 37/2007 y 1495/2011 que permiten la reutilización de la información pública en nuestro país. Esta política europea tiene una clara dimensión económica. Desea aportar valor tanto para la misma administración como para terceros. Se estima que el uso comercial y no comercial del *big data* en un mercado interno transparente favorecerá la economía, la competitividad, el empleo y la sociedad digital. En su origen, esta política parece estar más orientada a favorecer proyectos empresariales que a facilitar el acceso de los ciudadanos a sus datos. Debemos estar

* Autor para correspondencia.

Correo electrónico: joangenebadia@gmail.com (J. Gené Badia).

atentos a los cambios que introducirá el nuevo Reglamento Europeo de Protección de Datos que regulará este nuevo escenario que mezcla investigación de alto valor social con nuevos nichos de mercado.

Los países con buenas bases de datos sanitarias de carácter público han sido los primeros en aplicar la tecnología *big data* en el campo de la salud. Destacan el «care.data project» desarrollado en el Reino Unido y el proyecto VISC+, actualmente renombrado como PADRIS, de Catalunya.

Ambas iniciativas buscaban reutilizar la información clínica generada por la asistencia sanitaria para fines de investigación. Amparado en la legislación española, en 2015 el parlamento catalán otorgó a la agencia de evaluación del servicio de salud AQUAS la capacidad de facilitar datos sanitarios anonimizados de los pacientes a instituciones públicas y privadas. El proyecto, creado durante la época más dura de la crisis económica, tenía por objeto obtener una fuente adicional de financiación. Los impulsores calcularon que reportaría unos 25 millones de euros al sistema de salud. Contemplaba la posibilidad de explotar y comercializar datos a iniciativas con ánimo de lucro que podían combinar información clínica con otras bases de datos.

El entusiasmo de los impulsores chocó con la firme resistencia de una parte de la ciudadanía, algunos partidos de la oposición y ciertos movimientos profesionales que opinaban que el proyecto no aseguraba suficientemente la privacidad de los datos clínicos. Algunos grupos de investigación y comités de bioética emitieron informes al respecto, destacan uno crítico a cargo del Comité de Bioética y Derecho de la Universidad de Barcelona⁸ y otro más favorable redactado por el Comité de Bioética de Catalunya⁹.

Los críticos consideraban que la simple anonimización de los datos personales no garantizaba suficientemente la privacidad, precisamente porque la tecnología *big data* permite la reidentificación de las personas. La anonimización era una medida efectiva a finales de los años 90 y principio de los 2000 cuando se creó la legislación sobre protección de datos. La ley estaba pensada para las bases de datos tradicionales propias de aquel momento, pero actualmente con la tecnología *big data* tal legislación ha quedado obsoleta. Los avances tecnológicos corren más que la legislación. Sabemos que con los algoritmos adecuados y combinando distintas bases de datos podemos reidentificar individuos previamente anonimizados.

Surgen nuevos conflictos éticos: ¿cómo se evita, por ejemplo, que una vez entregados los datos para un proyecto de investigación específico, no sean posteriormente reutilizados para otros fines? ¿Están preparados los comités de ética de la investigación para evaluar adecuadamente proyectos que utilizan tecnología *big data*? ¿Cómo se garantiza el derecho individual a optar porque la información sanitaria personal facilitada para la atención clínica no sea posteriormente reutilizada con otros fines? Falta un verdadero debate social informado sobre la aplicación de estas tecnologías *big data* a la información sanitaria. Solo así los ciudadanos podrán conocer la existencia de estos proyectos, así como sus beneficios y riesgos. Una vez informados, deben disponer de un mecanismo ágil y transparente que les permita, si lo desean, ejercer el derecho de evitar que sus datos clínicos personales sean reutilizados para la investigación.

Mientras se producía este debate en Catalunya las elecciones aportaron un cambio de gobierno. El nuevo ejecutivo detuvo la licitación del proyecto VISC+ y procedió a su reorientación¹⁰. Lo renombró como PADRIS¹¹ y otorgó la responsabilidad exclusiva de su ejecución a la agencia AQUAS asesorada por un comité con miembros de la Agencia Catalana de Protección de Datos y el Comité de Bioética de Catalunya. Únicamente la administración sanitaria y los centros de investigación acreditados por la Generalitat tienen ahora acceso a los datos si consiguen que su proyecto de investigación sea aceptado por el comité ético. Aseguran la transparencia y la rendición de cuentas publicando los resultados de toda investigación que resulte autorizada. También afirman que el proceso de anonimización será público y auditado, y proponen mecanismos para retirar, si se desea, los datos sanitarios de los ficheros destinados a investigación.

Esta reorientación no resuelve todas las incógnitas, pero dificulta la explotación comercial y permite seguir aprendiendo en la gestión de la investigación del *big data* y en la preservación de la privacidad de la información. En nuestras consultas contribuimos a crear unas bases de datos que pueden ser extremadamente beneficiosas para nuestros pacientes al tiempo que suponen un reto a la privacidad y a la no discriminación. Como profesionales de la salud no podemos quedarnos solo fascinados por el avance científico, sino que también debemos velar por garantizar el respeto a la seguridad de la información personal. La inadecuación y complejidad de la legislación vigente y la corta vida de los proyectos desarrollados aconsejan extremar la prudencia.

Bibliografía

1. Harari YN. *Homo Deus*. Breve Historia del Mañana. Editorial Debate. 2016.
2. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *Proc Natl Acad Sci U S A*. 2015;112:1036–40. <http://dx.doi.org/10.1073/pnas.1418680112>.
3. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: Opportunities and policy implications. *Health Aff*. 2014;33:1115–22.
4. Kaisler S, Armour F, Espinosa JA, Money W. Big Data: Issues and Challenges Moving Forward. En: 46th Hawaii International Conference on System Sciences (HICSS); 2013. p. 995–1004. <http://dx.doi.org/10.1109/HICSS.2013.645>.
5. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. *Health Inf Sci Syst*. 2014;2:3.
6. Heitmueller A, Henderson S, Warburton W, Elmagarmid A, Pentland AS, Darzi A. Developing public policy to advance the use of big data in health care. *Health Aff*. 2014;33:1523–30.
7. Pérez G. Peligros del uso de los big data en la investigación en salud pública y en epidemiología. *Gac Sanit*. 2016;30:66–8.
8. Casado M, Buisán L, Llácer RM, editores. Documento sobre bioética y Big Data de salud: explotación y comercialización de los datos de los usuarios de la sanidad pública. Observatori de Bioètica I Dret. Catedra UNESCO de Biètica de la Universitat de Barcelona. Barcelona: Ed. UB; 2015 [consultado 17 Oct 2017]. Disponible en: http://www.bioeticayderecho.ub.edu/sites/default/files/doc_big-data.pdf
9. Comitè de Bioètica de Catalunya. Principis ètics I Directrius per a la Reutilització de la Informació del Sistema de Salut Català en la Recerca, la Innovació i l'avaluació Generalitat de Catalunya; 2015. [consultado 17 Oct 2017]. Disponible en:

- <http://comitebioetica.cat/wp-content/uploads/2015/05/3.-VISC-7-de-maig-2015.pdf>
10. Mocio¹⁵⁰/X del Parlament de Catalunya. [consultado 17 Oct 2017]. Disponible en: <http://www.parlament.cat/document/nom/10b421.pdf#page=24>
 11. Mocio⁴⁹/XI del Parlament de Catalunya. [consultado 17 Oct 2017]. Disponible en: <http://www.parlament.cat/document/bopc/172600.pdf>