Document sobre bioètica i Big Data de salut: explotació i comercialització de les dades dels usuaris de la sanitat pública

Documento sobre bioética y Big Data de salud: explotación y comercialización de los datos de los usuarios de la sanidad pública

Document on bioethics and Big Data: exploitation and commercialisation of user data in public health care

M. R. Llàcer, M. Casado i L. Buisan (coords.)

Organització de les Nacions Unides per a l'Educació, la Ciència i la Cultura

Càtedra UNESCO de Bioètica de la Universitat de Barcelona

Observatori de Bioètica i Dret

www.bioeticaidret.cat
www.bioeticayderecho.ub.edu
www.bioethicsandlaw.es

SUMARI

DOCUMENT SOBRE BIOÈTICA I BIG DATA DE SALUT: EXPLOTACIÓ I COMERCIALITZACIÓ DE LES DADES DELS USUARIS DE LA SANITAT PÚBLICA

PRESENTACIÓ

El Grup d'Opinió de l'Observatori de Bioètica i Dret, de la Universitat de Barcelona, es va constituir el 1996 dins l'Observatori de Bioètica i Dret, que té, entre altres objectius, analitzar amb una base científica i amb una metodologia interdisciplinària les implicacions ètiques, socials i jurídiques de les noves tecnologies i els problemes biotecnològics i biomèdics, a fi d'intervenir en el diàleg entre la universitat i la societat mitjançant la transmissió del coneixement científic i tècnic, i aportant els arguments necessaris per a contribuir al debat social informat. Amb aquesta finalitat, el grup d'opinió ha elaborat ja vint-i-dos documents[1] sobre temes d'actualitat i sobre els quals no hi ha una opinió unànime, ni en la societat ni en les diverses comunitats científiques implicades; això ha fet necessari identificar els problemes, contrastar els arguments i proposar recomanacions de consens.

En aquesta ocasió, el grup fa públic el document d'opinió *Bioètica i Big Data de salut: explotació i comercialització de les dades dels usuaris de la sanitat pública,* a fi de cridar l'atenció sobre la necessitat de crear una cultura de la privacitat respecte de les dades personals, que han esdevingut elements o mecanismes de control en una societat informatitzada, per la qual cosa cal ser conscients de per què i per a què aquestes dades han de ser protegides. Aquest document analitza, des de la perspectiva bioètica, els problemes de l'explotació i la comercialització de dades dels usuaris de la sanitat pública. Prenent com a punt de partida el reconeixement del principi d'autonomia de les persones, el document posa de manifest que la implementació de les tecnologies Big Data en l'àmbit sanitari, associada a una eventual comercialització d'aquestes dades, impacta directament en el nostre sistema sanitari i investigador —fonamentat en els principis d'igualtat i no-discriminació— i afecta de ple a l'àmbit privat dels ciutadans.

El motiu immediat d'aquest document han estat els problemes detectats en el projecte VISC+ (formalment anomenat Més Valor a la Informació de Salut de Catalunya) tant en relació amb possibles vulneracions dels drets dels ciutadans com amb la manca de transparència i de debat públic informat en una qüestió de tanta importància com és el tràfic de dades personals, reutilitzades amb finalitats diferents de l'atenció mèdica que puguin rebre directa-

[1] Tots els documents del Grup d'Opinió de l'Observatori de Bioètica i Dret són accessibles en format PDF i en obert a: www.bioeticaidret.cat/documents (versió en català, espanyol i anglès) [consultat el 27 de gener de 2015].

ment els ciutadans. Els arguments que aquí oferim no solament fan referència a l'esmentat projecte, sinó que tenen un abast major, perquè tenen a veure amb: *a*) la validesa de les tècniques d'anonimització en els conjunts de dades (*datasets*); *b*) la necessitat de redefinir el concepte de dades personals, tenint en compte la possibilitat actual de reidentificar les persones, i *c*) l'impacte d'aquests dos aspectes en els mercats emergents de Big Data, *data marketplaces* i *digital marketing*.

Creiem que cal prendre mesures que garanteixin l'exercici dels drets i les decisions lliures i informades de totes les persones implicades. Pretenem obrir el debat sobre aquestes qüestions fent propostes que permetin afrontar el canvi de paradigma que impliquen aquestes noves tecnologies de la informació, car en una societat democràtica les decisions de l'Administració no han de ser imposades al ciutadà sense una informació prèvia, veraç i transparent sobre el seu abast.

Aquest document ha estat coordinat per les Dres. Maria Rosa Llàcer, María Casado i Lídia Buisan i ha estat elaborat pel Grup d'Opinió de l'Observatori de Bioètica i Dret de la Universitat de Barcelona, el Grup de Recerca Consolidat «Bioètica, Dret i Societat» de la Generalitat de Catalunya, amb la col·laboració del Grup de Recerca Consolidat «Dret Privat, Consum i Noves Tecnologies» (GREDINT) de la Generalitat de Catalunya. En la seva elaboració hi han participat, a més, les persones els noms i perfils professionals de les quals s'inclouen al final del document.

CONSIDERACIONS GENERALS

Els reptes del Big Data i l'anonimització

L'expressió *Big Data* és un terme que fa referència al tractament massiu de dades per mitjà d'algoritmes matemàtics a fi de generar correlacions entre elles, predir tendències i prendre decisions. Les tecnologies Big Data constitueixen un paradigma nou que, a més, implica canvis organitzatius importants tant en les empreses com en l'Administració. En l'actualitat, els objectius empresarials no són ja només la millora dels processos, sinó la gestió de les dades. Estem assistint a una fase de transició cap a la *datificació* i la *monetització*, fet que comporta extreure un valor nou de les dades i rendibilitzar-les, tant en l'àmbit privat com en el públic o bé en una combinació de tots dos. Es tracta d'una tendència que s'insereix en el marc d'una indústria creixent basada en el coneixement adquirit mitjançant la reutilització i l'explotació de les dades, qüestió que cal tenir en compte a fi de contextualitzar el debat i entendre millor aquest canvi de model. Això no obstant, l'aposta per la innovació no ha de fer oblidar els aspectes ètics i els drets fonamentals de les persones, ni la protecció dels ciutadans en el context d'aquests nous avenços de les tecnologies. Es tracta de plantejar l'anàlisi d'aquesta situació amb la finalitat de proposar un nivell de protecció fort que suposi, per això mateix, un grau més avançat d'innovació en aquest àmbit.

És molt important assenyalar que, fins ara, la premissa de l'*anonimització* de les dades ha estat la garantia que permetia respectar les regulacions de protecció de dades personals existents, amb el benentès que en ser anonimitzada la dada personal passa a ser simplement una dada, perdent així la protecció de la normativa de protecció de dades personals, normativa que pretén ser rigorosa, tant a la Unió Europea com a l'Estat espanyol, però que amb els avenços de les tecnologies informàtiques després de gairebé vint anys ha esdevingut en bona part obsoleta. El problema rau en el fet que, actualment, aquesta anonimització és palesament il·lusòria, perquè mitjançant tècniques d'enginyeria informàtica hom pot tornar a connectar les dades amb la persona font.[2] Tant la desanonimització de les dades com la reidentificació

[2] Art. 29 Working Party Opinion 06/2013 on open data and public sector information ('PSI') reuse (1021/00/EN WP207); Art. 29 Working Party Opinion 05/2014, on Anonymisation Techniques (0829/14/EN WP216). Vegeu: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

de persones són possibles si es disposa de la competència tècnica i dels mitjans necessaris per a fer-ho, per la qual cosa el debat es trasllada a un territori més tècnic i objectiu que proporciona informació i arguments que afecten directament la cada vegada més estesa indústria de venda de dades. N'hi ha prou amb saber que la reidentificació es pot fer tenint en compte el valor especial que poden adquirir determinades dades que fins ara s'han considerat no personals; per exemple, avui dia és prou evident que amb el codi postal, la data de naixement i el sexe ja és possible reidentificar la major part de les persones d'un *dataset*.[3] De manera semblant a com les nostres empremtes digitals ens identifiquen de manera unívoca, el mateix passa amb determinades tipologies de dades. La polèmica que hi ha darrere no és gens banal: què és una dada personal i com en podem garantir la protecció?[4] Com es pot evitar que a partir d'un conjunt de dades no personals es pugui identificar una persona?

Volem remarcar especialment aquest punt perquè el negoci de «posar en valor» les dades depèn precisament del concepte d'anonimització esmentat, ja que seria precisament aquest el que permetria complir amb les regulacions de protecció de dades personals. El debat sobre l'anonimització, tot i que ja té una certa història, no ha fet més que començar i és ben lluny d'estar acabat. En la nostra opinió, aquesta discussió és crucial en el segle XXI i no està tenint, ni de bon tros, la presència que li escauria en els diferents fòrums que hi tenen a veure (legals, ètics, tècnics, empresarials, governamentals) i en els quals caldria endegar el debat oportú perquè sigui compresa, primer, i resolta o si més no gestionada, després.

Com s'ha dit, actualment les evidències tècniques ja ens mostren que és possible reidentificar persones concretes a partir de les dades d'un *dataset* al

---

Narayanan, Arvin; Felten, Edward W. «No silver bullet: De-identification still doesn't work», 2014. Vegeu: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.

Narayanan, Arvin; Shmatikov, Vitaly. «Robust de-anonymization of large sparse datasets», *Security and Privacy*, IEEE Symposium on IEEE, 2008, pp. 111-125.

De Montjoye, Yves-Alexandre, *et al*. «Unique in the Crowd: The privacy bounds of human mobility». *Scientific Reports*, 2013, vol. 3.

Ohm, Paul. «Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization», *UCLA Law Review*, 2009-2010, p. 731.

[3] Sweeney, L. *Simple demographics often identify people uniquely*. Carnegie Mellon University, editor. Data Privacy Working Paper 3. 2000. http://dataprivacylab.org/projects/identifiability/paper1.pdf [consultat el 27 de gener de 2015].

[4] Schwartz, Paul M.; Solove, D. «The PII Problem: Privacy and a New Concept of Personally Identifiable Information», *New York University Law Review*, vol. 86, 2011, p. 1814.

qual s'hagin aplicat prèviament tècniques d'anonimització (o de desidentificació). Una persona, o bé una empresa, poden aconseguir la reidentificació esmentada si hi tenen interès (per motius econòmics, empresarials, delictius...), i si tenen també els coneixements i els mitjans tecnològics per a fer-ho (per exemple, si disposen de les dades sanitàries d'un hospital —encara que no continguin dades personals— i d'accés a les dades personals d'un altre *dataset*, com ara un cens). Resulta evident que, en el cas de les dades de salut, no és difícil trobar un suposat «adversari» amb la motivació i els recursos per a fer-ho, i és escaient, per tant, qüestionar la validesa de les iniciatives de bescanvi de dades sensibles fonamentades en tècniques d'anonimització. En l'àmbit jurídic, l'incert recurs a l'«anonimització», entesa com una solució definitiva però inevitablement en crisi, es recolza en la normativa actual de protecció de dades, que prové d'una directiva europea de l'any 1995 —per tant, molt anterior al fenomen del Big Data— i que es recull en la Llei 37/2007, del 16 de novembre, sobre la reutilització de la informació del sector públic. Però si el concepte mateix d'anonimització esdevé incert, cal trobar un fonament que legitimi l'anàlisi de dades personals de salut a gran escala. Si no és així, aleshores s'està obrint la porta a usos no desitjats d'aquestes dades, car en haver donat anteriorment el titular de les dades el seu consentiment per a determinades accions en l'àmbit sanitari i de recerca, en realitat en perd el control i queda desprotegit sense que sàpiga —perquè té una concepció equivocada de la protecció de dades i del secret professional— que les seves dades poden haver sigut utilitzades o cedides per a altres fins que no són ni desitjats ni efectivament consentits.

Aquest document no pretén rebutjar, sense més ni més, aquest nou model de negoci que centra l'atenció del mercat i en el qual, a més, ja estem immersos, sinó que vol alertar, tant als ciutadans com als poders públics que regulen i controlen l'activitat en l'àmbit sanitari i de recerca, dels riscos que comporta. En el nostre entorn no està prou consolidada una consciència social de la importància de protegir les dades en relació amb el dret fonamental a la intimitat i a la no-discriminació. No tenim una cultura de la privacitat que ens permeti comprendre de quina manera ens pot afectar que una empresa acumuli i faci rendible la nostra informació, i que disposi d'un instrument de poder en base al qual pugui prendre decisions que ens afectin.[5] N'és un exemple el fet que l'anàlisi massiva de dades es pot fer servir per a descobrir efectes secundaris de medicaments, però també per a generar per-

[5] Cohen, Julie. «What Privacy is for», *Harvard Law Review,* 126, 2012-2013, p. 1904.

fils de risc —i que els propis afectats poden no conèixer— que es podrien utilitzar per a «justificar» la denegació d'una assegurança.

Es fa evident, en conseqüència, la urgència d'un debat que posi en relleu la vulnerabilitat de les persones davant el risc de discriminació generat per perfils i patrons de conducta establerts amb finalitats que la persona afectada no pot controlar, i també sobre l'adaptació de les lleis als reptes ètics i socials que les Big Data plantegen. Aquest és, precisament, el nucli del conflicte que cal obrir al debat públic amb implicació de la ciutadania a fi de crear una cultura de la privacitat que vagi d'acord amb l'actualitat i amb les noves realitats.[6]

Sobre el projecte VISC+

Un exemple de reutilització de dades que, des del nostre punt de vista, resulta molt qüestionable és el projecte VISC+, impulsat per l'Agència de Qualitat i Avaluació Sanitàries de Catalunya (AQuAS), que té com a objectiu —segons esmenten els seus promotors— posar la informació sanitària a disposició dels ciutadans, les empreses i la recerca per a millorar els serveis de salut i la investigació i per a «posar en valor» el coneixement.

L'esmentat projecte es nodreix de les diferents bases de dades que ja existeixen en el sistema sanitari: el Sistema d'Informació per al Desenvolupament de la Investigació en Atenció Primària (SIDIAB) i, de manera especial, la Història Clínica Compartida a Catalunya (HC3), que recull les dades assistencials i de consum farmacèutic, juntament amb altres informacions rellevants com són la identificació i la situació sociosanitària de cada ciutadà atès per la sanitat pública; a més, la HC3 conté informació de les proves analítiques i diagnòstiques que inclouen paràmetres metabòlics i bioquímics, així com dades de diagnòstic genètic que identifiquen les persones portadores de malalties genètiques hereditàries o bé que mostren riscos i susceptibilitats de patir malalties més complexes. Aquestes bases de dades fan que existeixin «fitxers d'usuaris» dels quals és responsable el Departament de Salut de la Generalitat de Catalunya. Tot i que els macrofitxers de dades estan protegits per la normativa ja existent —en especial per la Llei Orgànica de Protecció de Dades (LOPD) i el Reglament que la desenvolupa—, aquesta regulació ha esdevingut del tot inadequada en el marc de la nova realitat dels

[6] Richards, N. M.; King, J. H. «Big data ethics», *Wake Forest Law Review*, 49, 2014, pp. 393-432.

Big Data, com s'ha esmentat abans, i no garanteix de cap manera que no es produeixin usos indeguts i discriminatoris.[7]

La HC3 té els objectius explícits següents: a) millorar l'atenció de la salut dels ciutadans mitjançant una eina que faciliti la feina dels professionals sanitaris respecte als malalts als quals han d'atendre; b) propiciar un nou model assistencial en permetre l'accés i la consulta de forma immediata, segura i confidencial, de la informació rellevant disponible sobre els usuaris. Com és obvi, les dades que conté la HC3 són extremament sensibles, la recollida i el tractament de les quals es justifica en la seva eficàcia per a proporcionar una assistència de qualitat, no sols en el centre que habitualment atén l'usuari sinó en tota la xarxa assistencial pública de Catalunya,[8] perquè la HC3 permet l'accés de manera organitzada, i atenent a criteris de seguretat i confidencialitat, a les històries clíniques de la xarxa assistencial. Aquesta eina ha d'oferir beneficis tant a la ciutadania, com als professionals sanitaris, com al mateix sistema de salut. Per aquesta raó, el ciutadà té dret a saber qui pot accedir a les seves dades personals i amb quina finalitat; i també té dret a exigir responsabilitats si creu que se n'està fent un ús indegut o distint d'aquell al qual en el seu moment va consentir. Quan es va dur a la pràctica la HC3, ni es va informar suficientment els ciutadans d'aquesta recollida massiva de dades amb finalitat assistencial, ni es va indicar en cap moment que aquestes mateixes dades podrien ser reutilitzades amb altres finalitats, fins i tot comercials. De cap manera es pot considerar que la cessió de les dades per a finalitats no assistencials sigui el «preu» de la gratuïtat de l'assistència sanitària, ja que si s'exigís algun tipus de contraprestació l'assistència deixaria de ser gratuïta.

La informació que conté la HC3, tot i ser recollida i estructurada pels professionals assistencials, fa referència a les dades de salut del malalt i, per tant, aquestes li pertanyen, per la qual cosa les entitats assistencials reben peticions relacionades amb l'exercici dels drets que la normativa sobre protec-

---

[7] Com el grup d'opinió ja va advertir en el *Document sobre proves genètiques de filiació*, Barcelona: Signo, 2006. Disponible en format PDF a: www.publicacions.ub.edu/refs/observatoriBioEticaDret/documents/07899.pdf.

[8] Existeix un projecte d'un abast territorial major, desenvolupat en dotze països de la Unió Europea, en el marc de www.epSOS.eu i del qual formen part Alemanya, Àustria, Txèquia, Dinamarca, Eslovàquia, Espanya, França, Grècia, Holanda, Itàlia, el Regne Unit i Suïssa. De l'Estat espanyol hi participen tres comunitats autònomes: Andalusia, Castella-La Manxa i Catalunya, dins del Plan Avanza per a la modernització dels serveis de les Administracions públiques.

ció de dades vigent reconeix a la ciutadania: drets d'accés, de rectificació, de cancel·lació i d'oposició (els anomenats drets ARCO). La finalitat d'aquest conjunt de drets és impedir un tractament il·lícit i lesiu per a la dignitat i el dret de l'afectat (*habeas data*), alhora que garantir l'exercici del dret més general a la intimitat. Els usos de les dades recollides en la HC3 s'han de limitar a l'assistència (juntament amb les finalitats científiques —en epidemiologia, investigació i docència, o bé dirigides a la millora dels serveis públics— que la normativa actual ja autoritza) i és absolutament necessari establir garanties reals que evitin el tràfic de dades i qualsevol ús indegut per part d'empreses de l'àmbit de la salut (assegurances mèdiques, corporacions farmacèutiques, entitats financeres, i altres). Per això, el projecte VISC+, tal com en aquests moments està previst que es dugui a terme, genera dubtes importants, tant de caràcter bioètic com estrictament jurídic, que convé analitzar amb detall i debatre, a fi de prevenir-ne possibles usos discriminatoris.

Problemes rellevants del projecte VISC+

1. Denominació equívoca del projecte

La mateixa denominació del projecte és equívoca i no s'adequa al principi general de lleialtat en la recollida i tractament de dades, perquè indueix a pensar que el projecte ajuda a millorar les condicions de vida i la salut dels ciutadans. Els usuaris a qui es demani el consentiment perquè les seves dades personals de salut siguin tractades en el marc d'un projecte anomenat VISC+ poden pensar, erròniament, que col·laboren en un programa que l'única cosa que pot aportar-los són beneficis. Això pot ser fàcilment relacionat amb una pràctica deslleial, entesa com una conducta contrària a la bona fe objectiva que distorsiona la capacitat d'escollir amb ple coneixement de causa i que indueix a facilitar unes dades que altrament no s'haurien proporcionat. La lleialtat és un valor fonamental en el marc de la LOPD, ja que la recollida i el tractament de dades per mitjans fraudulents o deslleials està prohibida expressament i afecta de manera directa el principi de qualitat de les dades.

2. Limitació de les finalitats en el tractament de les dades

Com bé es diu en l'informe elaborat pel *Grup Europeu d'Ètica de les Ciències i de les Noves Tecnologies* (GEE), de la Comissió Europea, en la recollida i el

tractament de dades de caràcter personal les entitats públiques i les privades han de fonamentar la seva activitat en el principi de «limitació de la finalitat»;[9] és a dir, que aquest tipus de dades no haurien de ser recollides ni tractades per a qualsevol ús, sinó només amb objectius específics i legítims. Cal, a més, que les dades no estiguin per defecte a disposició de «qui les vulgui utilitzar» i que els ciutadans tinguin mecanismes efectius per a controlar i modificar les informacions que els concerneixin i que estiguin dipositades en les esmentades entitats. Insisteix també l'informe que la possible cessió de dades amb finalitats comercials ha de fer-se només amb el consentiment explícit de les persones afectades, i que les entitats privades han d'indicar el tipus de dades que preveuen tractar i amb quin objectiu, durant quant de temps, i si tenen intenció de relacionar o connectar aquestes dades amb altres procedents de diferents fonts.

Resulta especialment important, en el marc del projecte que aquí s'analitza, la gradualitat de la protecció de les dades en funció de la finalitat de l'ús, distingint acuradament les finalitats sanitària, epidemiològica i d'investigació i docència —que ja estan emparades per la legislació— de les finalitats privades, a les quals cal exigir el nivell de protecció més elevat. Ara bé, el projecte VISC+ equipara tractaments que tenen finalitats totalment diferents i aquesta confusió afecta la legitimació per tractar dades sanitàries personals, que són dades especialment sensibles i que, en conseqüència, estan sotmeses a una protecció especial.

Com ja s'ha dit, la legislació empara l'ús de dades dels usuaris de salut per a dur a terme l'assistència sanitària i per a la recerca i la millora dels serveis públics. Però si es vol anar més enllà i facilitar la utilització d'aquestes dades amb finalitats ni previstes ni autoritzades —com són els interessos comercials d'empreses privades els productes de les quals depenen de la recerca, a més d'altres factors— cal un debat social previ sobre la concurrència dels interessos públics i privats en l'àmbit de la recerca a fi definir els límits ètics i el nivell de protecció de què disposarà el ciutadà quan empreses amb interessos privats facin el tractament de dades de salut. L'*empowerment* del ciutadà es basteix amb informació adequada, clara i veraç sobre l'ús de les seves dades, a més de reconèixer-li la facultat de controlar-ne el tractament, ja sigui consentint-hi o bé oposant-s'hi.

---

[9] European Group on Ethics in Science and New Technologies. *Ethics of Security and Surveillance Technologies*. Opinion n. 28, 20 de maig de 2014. Vegeu: http://ec.europa.eu/archives/bepa/european-group-ethics/docs/publications/ege_opinion_28_ethics_security_surveillance_technologies.pdf.