



UNIVERSITAT DE
BARCELONA



Revista de Bioética y Derecho

Perspectivas Bioéticas

www.bioeticayderecho.ub.edu - ISSN 1886-5887

DOSSIER BIG DATA

Obert per defecte, tancat quan sigui necessari: el cas de les dades de la recerca

Open by default, closed when necessary: the case for data in research

Abierto por defecto, cerrado cuando sea necesario: el caso de los datos de investigación

IGNASI LABASTIDA I JUAN*

OBSERVATORI DE BIOÈTICA I DRET DE LA UNIVERSITAT DE BARCELONA

La Revista de Bioética y Derecho se creó en 2004 a iniciativa del Observatorio de Bioética y Derecho (OBD), con el soporte del Máster en Bioética y Derecho de la Universidad de Barcelona: www.bioeticayderecho.ub.edu/master. En 2016 la revista Perspectivas Bioéticas del Programa de Bioética de la Facultad Latinoamericana de Ciencias Sociales (FLACSO) se ha incorporado a la Revista de Bioética y Derecho.

Esta es una revista electrónica de acceso abierto, lo que significa que todo el contenido es de libre acceso sin coste alguno para el usuario o su institución. Los usuarios pueden leer, descargar, copiar, distribuir, imprimir o enlazar los textos completos de los artículos en esta revista sin pedir permiso previo del editor o del autor, siempre que no medie lucro en dichas operaciones y siempre que se citen las fuentes. Esto está de acuerdo con la definición BOAI de acceso abierto.

* Ignasi Labastida i Juan. Oficina de Difusión del Conocimiento del Centro de Recursos para el Aprendizaje y la Investigación (CRAI), Universitat de Barcelona. Espanya. E-mail: ilabastida@ub.edu.

Resum

Les darreres decisions polítiques dels agents finançadors de la recerca han estat encaminades a fomentar la difusió dels resultats en obert, és a dir a oferir-los a tothom, sense cap mena de restricció, i permetent-ne la reutilització. Aquestes polítiques que van començar aplicant-se a les publicacions, s'han estès a les dades, un bé molt preuat entre els investigadors als quals no els és fàcil desprendre-se'n. El fet d'haver de compartir les dades ha obert un debat sobre quines dades cal compartir i com s'ha de fer-ho. A més, han sorgit moltes reticències algunes de les quals s'han escudat en el fet que hi ha dades que difícilment es podran oferir mai en obert atesa la seva naturalesa: dades personals o sensibles. En aquest text volem aclarir alguns aspectes d'aquestes polítiques encaminades a obrir les dades sempre que sigui possible i a tancar-les quan calgui.

Paraules clau: dades; recerca; accés obert.

Abstract

The latest decisions on policies from research funders have been aimed at promoting open dissemination of research outputs, available to everyone, without any restriction, and allowing their reuse. These policies, which began to be applied to publications, have been spread to research data, a well appreciated good among researchers who do not easily get rid of it. The requirement of sharing data has opened a debate on what to share and how to do it. In addition, there have been reluctances, some of them shielded from the fact that some data would never be open due to its nature: personal or sensitive data. In this text we want to clarify some aspects of these policies addressed to open data whenever it is possible and close them when it is necessary.

Keywords: data; research; open access.

Resumen

Las últimas decisiones políticas de los agentes financiadores de la investigación han sido encaminadas a fomentar la difusión de los resultados en abierto, es decir a ofrecerlos a todo el mundo, sin ningún tipo de restricción, y permitiendo la reutilización. Estas políticas que empezaron aplicándose a las publicaciones, se han extendido a los datos, un bien muy preciado entre los investigadores a los cuales no les es fácil desprenderse. El hecho de tener que compartir los datos ha abierto un debate sobre qué datos hay que compartir y cómo se tiene que hacer. Además, han surgido muchas reticencias algunas de las cuales se han escudado en el hecho que hay datos que difícilmente se podrán ofrecer nunca en abierto dada su naturaleza: datos personales o sensibles. En este texto queremos aclarar algunos aspectos de estas políticas encaminadas a abrir los datos siempre que sea posible y a cerrarlos cuando sea necesario.

Palabras clave: datos; investigación; acceso abierto.

Fa més de quinze anys que es va publicar el manifest de Budapest¹ que advocava per posar a disposició del públic la literatura que els acadèmics difonen sense cap expectativa de cobrament directe, és a dir els articles revisats per parells i publicats en revistes científiques, així com els textos sense revisar que posen en línia per rebre'n comentaris o per advertir els col·legues de les seves troballes. L'accés a totes aquestes obres hauria de ser gratuït i, a més, podrien ser utilitzats sens cap mena de restricció a canvi de reconèixer-ne adequadament l'autoria i de mantenir-ne la integritat. Amb aquest manifest s'establien els fonaments del que s'ha anomenat *moviment de l'accés obert* i que ha comportat un sacseig en la manera com es difonen actualment els resultats de la recerca².

L'any 2012, per celebrar el desè aniversari del manifest de Budapest es va elaborar un nou document on es proposaven unes noves recomanacions per a la propera dècada. Entre aquestes recomanacions, se citaven les dades complementàries a les publicacions com un dels resultats que caldria també posar en obert, és a dir a disposició de tothom, de manera gratuïta i per ser reutilitzades sense restriccions³.

El moviment de l'accés obert s'ha anat consolidant al llarg d'aquests anys gràcies a l'aposta de les agències de finançament, tant públiques com privades, que han adoptat polítiques encaminades a facilitar l'accés als resultats de la investigació que han finançat. Aquest accés, normalment és sense restriccions i cada cop més immediat. Els primers resultats als quals es tenia accés eren les publicacions però cada cop més es pot accedir al codi, les dades o fins i tot als quaderns de laboratori. Aquesta nova manera de fer més visible i accessible tot el cicle de la recerca rep el nom de *ciència oberta*.

En els darrers anys la Comissió Europea ha volgut impulsar la ciència oberta, primer fent una consulta pública⁴, després impulsant programes i accions per reconèixer aquesta manera oberta de desenvolupar una investigació, i finalment, elaborant, juntament amb el govern neerlandès, un document cridant a l'acció⁵. Tanmateix, encara hi ha molts reptes en aquest àmbit i cal redissenyar els processos d'avaluació.

Tot i l'impuls per afavorir l'obertura de la recerca, cal vèncer oposicions que venen de tot tipus de llocs. Fa un anys els principals opositors del moviment de l'accés obert eren les editorials científiques que veien perillar el seu model. Però les editorials han anat adoptant decisions per treure'n benefici, especialment amb l'ofertament del model híbrid que permet que un investigador ofereixi la seva publicació en accés obert en una revista d'accés restringit després de pagar una quantitat gens menyspreable i que en molts cops supera el que algunes revistes d'accés obert demanen per publicar un article. Aquest fet és fàcilment detectable en els informes anuals que elaboren les agències de finançament, com per exemple Wellcome Trust⁶.

Però quan parlem de compartir dades no són les editorials les que principalment s'hi oposen, Algunes fins i tot hi donen suport i han adoptat polítiques requerint que les dades estiguin accessibles abans d'iniciar el procés de revisió⁷.

Les principals reticències provenen dels mateixos investigadors. En una enquesta realitzada l'any 2015 a Àustria⁸, els investigadors senyalaven les principals causes per no compartir les dades: l'esforç necessari per polir les dades, els possibles abusos en utilitzar-les, les ambigüitats legals, o la pressió de la competitivitat entre investigadors. També cal destacar l'editorial apareguda a principis de 2016 al *New England Journal of Medicine* titllant de paràsits als investigadors que aprofitessin les dades publicades per a la seva recerca⁹. Aquesta editorial va provocar la resposta de molts investigadors defensant que el progrés de la ciència sempre s'ha basat en els resultats obtinguts prèviament. Un investigador britànic afirmava «Crec que compartir completament els resultat de la recerca, incloent-hi dades i codi, és essencial per accelerar la investigació i de fet jo me n'he aprofitat en la meva recerca»¹⁰.

Així doncs, veiem que encara hi ha molts dubtes, incerteses i inseguretats per compartir obertament les dades associades a una activitat investigadora. Cal que els investigadors tinguin clar què se'ls està demanant per part dels finançadors quan se'ls requereix compartir les dades per defecte i tancar-les quan sigui necessari, tal i com estipula, per exemple, l'actual programa d'ajuts a la recerca de la Comissió Europea.

1. Les polítiques que obren

Al juliol de 2012 la Comissió Europea feia pública una comunicació¹¹ on marcava les línies futures dels propers programes marc de finançament de la recerca i on destacava una clara aposta cap a la difusió en obert dels resultats. Aquesta aposta ve quedar reflectida en el programa Horitzó 2020 on no hi havia dubte respecte a les publicacions: totes havien de ser accessibles al públic en un termini màxim de sis o dotze mesos, segons la disciplina, després de la publicació formal. A l'inici d'aquest programa també es va posar en marxa l'anomenat pilot de dades obertes¹² adreçat a un nombre limitat de convocatòries. A partir de l'any 2017, el pilot de dades es va estendre a tots els projectes finançats en el marc del programa Horitzó 2020¹³ i per tant l'afectació entre els investigadors és molt més gran.

El que estipula aquesta política respecte els dades de la recerca és que, per defecte, les dades que suporten qualsevol resultat publicat han d'estar accessibles en obert facilitant-ne la reutilització. A més, cal que tots els projectes lliurin un pla de gestió de dades en un termini no

superior als 6 mesos després de l'inici del projecte. Aquesta nova política ha generat una sèrie de dubtes entre els investigadors sobre quines dades han de compartir, on les han de publicar, quan ho han de fer, qui hi podrà accedir i qui les podrà reutilitzar. Cal aclarir que el pilot estableix una sèrie d'exempcions com per exemple la utilització de dades personals, dades confidencials o dades relacionades amb la seguretat, entre altres. En aquests casos cal tancar-ne l'accés i fer-ho constar en el corresponent pla de gestió. L'objectiu de la Comissió és, principalment, millorar la gestió de les dades per part dels investigadors i facilitar la reproducibilitat de la recerca. També cal dir que no es pretén obrir totes les dades amb les quals es treballa al llarg d'una investigació sinó les dades amb les quals es fonamenten els resultats finals publicats.

Per alguns investigadors, aquesta política europea ha estat la primera vegada que se'ls ha demanat compartir les dades més enllà de l'intercanvi habitual entre col·legues o grups de recerca. Tanmateix, les polítiques sobre les dades ja existeixen en alguns països des de fa un any, com per exemple al Regne Unit. No només els finançadors han establert polítiques sinó que les institucions on es desenvolupa la recerca també tenen polítiques al respecte on es demana als investigadors que arxivin les dades en infraestructures determinades per poder preservar-les correctament durant un període de temps establert. En aquestes polítiques institucionals s'estableixen drets i deures a tots els nivells de la comunitat.

2. Com fer un pla de gestió de dades

La majoria de polítiques sobre dades de la recerca establertes per una institució que finança activitats de recerca requereixen que els beneficiaris elaborin un pla de gestió de dades. És el cas de la Comissió Europea¹⁴ però també d'altres institucions, fundacions o agències^{15, 16}. Tot i que podríem trobar alguna diferència entre els diferents models de pla de gestió que es demanen, podem establir uns requeriments comuns a l'hora de l'elaborar el pla. Primer cal descriure quines dades s'utilitzaran, ja siguin dades existents o dades que s'obtindran al llarg de la recerca. En aquesta descripció cal incloure els formats que s'utilitzaran i si seguiran els estàndards existents. Cada disciplina o àmbit de recerca pot tenir uns formats i uns estàndards específics i és per aquesta raó que no n'hi ha uns d'establerts comuns per a tots els projectes. En el pla també cal indicar qui tindrà accés a les dades i quan hi podrà accedir. Les polítiques encaminades a obrir els resultats de la recerca requereixen que en algun moment les dades es facin públiques però es poden establir diferents terminis i modalitats d'accés al llarg del procés de la investigació, per exemple facilitant només accés als participants del projecte en una primera instància. En aquest punt és on cal emfatitzar la sensibilitat de les dades, indicant-hi, per exemple, que l'accés a

determinades dades quedarà restringit sempre atesa la seva naturalesa o procedència, per exemple en el cas de dades personals o dades confidencials. Finalment cal establir on es dipositaran les dades al llarg del projecte. Aquí cal tenir molt en compte la vulnerabilitat de determinats llocs d'emmagatzematge quan estem tractant amb dades sensibles. Els investigadors haurien de vetllar per mantenir les dades en llocs on el risc de pèrdua o mal ús fos mínim. Al final del projecte caldrà establir on s'arxivaran i es preservaran les dades per a una possible reutilització un cop finalitzat i durant quan de temps estaran disponibles. Si s'ofereix accés al públic, com demanen algunes polítiques, caldrà establir el grau de reutilització permès, normalment mitjançant alguna llicència. De vegades, aquestes llicències no s'apliquen directament a les dades sinó a l'estructura de les bases de dades. És a dir, són llicències que permetem que es puguin extreure i reutilitzar les dades. En els plans de gestió també s'han d'incloure les despeses que comporta aquesta gestió, la responsabilitat de les persones implicades i dels diferents processos implicats així com els possibles drets de propietat intel·lectual que hi pugui haver.

El pla de gestió és un document dinàmic que pot evolucionar al llarg del temps atesos els canvis que es poden produir al llarg del procés de la recerca.

En el cas dels projectes finançats en el marc del programa Horitzó 2020 el model de pla de gestió ha anat modificant-se. En l'inici del pilot es demanava una descripció detallada de cada conjunt de dades fent que el pla fos costós d'elaborar i podria ocupar pàgines i pàgines. Amb la generalització del pilot ara el que es demana és una descripció general de totes les dades amb les quals es treballarà¹⁴. No obstant això, ara es demana què és farà per tal que les dades siguin FAIR. Aquest acrònim prové de quatre termes en anglès: Findable, Accessible, Interoperable i Reusable, que analitzarem en la propera secció.

3. Els principis FAIR

Al març de 2016 es publicava un article a la revista *Scientific Data*¹⁷ demanant millores en les infraestructures per facilitar la reutilització de les dades. Per assolir aquesta reutilització, els autors proposaven establir uns principis que caldria seguir. Aquests principis van rebre el nom de FAIR, seguint l'acrònim en anglès de trobables (Findable), accessibles (Accessible), interoperables (Interoperable) i reutilitzables (Reusable). Així doncs, les dades de la recerca s'haurien de poder trobar mitjançant identificadors permanents, com per exemple els DOI. Haurien de ser accessibles no només a humans sinó també a màquines mitjançant protocols oberts. Haurien de seguir estàndards o ontologies reconegudes en la disciplina científica

corresponent. I finalment, caldria facilitar-ne la reutilització sempre que se'n reconeguis la procedència. A més aquest principis s'estenen a les metadades que descriuen les dades i que són indispensables que s'hi trobin associades per tal de fer que les dades siguin trobables i accessibles. Quan l'accés a les dades no pugui ser públic per temes, per exemple, de confidencialitat, intimitat o seguretat, és fonamental que les metadades segueixin els principis FAIR. D'aquesta manera sabrem que les dades existeixen, han estat recollides i compilades, i potser s'han establert uns mecanismes o procediments per accedir-hi de forma controlada.

Aquests principis han estat adoptats per la Comissió Europea i actualment es demana als investigadors que reben ajuts per a projectes de recerca que expliquin en el pla de gestió corresponent què farem per tal que les dades de la investigació siguin FAIR. En aquest cas cal indicar si es faran servir identificadors, on s'ubicaran, quins protocols d'accés se seguiran i quins formats o estàndards s'utilitzaran. Pel que fa a la reutilització es recomana l'ús de llicències per a continguts oberts, com per exemple les que ofereix l'organització Creative Commons, i en especial les llicències que pràcticament no imposen cap restricció sempre que se'n reconegui l'autoria i la procedència, com la CC BY i la CC0. Cal dir que en molts casos les llicències no s'apliquen a les dades perquè són continguts no protegibles per propietat intel·lectual però sí que s'aplica a l'estructura de la base de dades que les conté¹⁸. D'aquesta manera se'n pot extreure tot el contingut o una part substancial per ser utilitzat sense haver de demanar permís al fabricant de la base dades, tal com estableixen les lleis de propietat intel·lectual vigents a la Unió Europea.

4. Les dades tancades quan sigui necessari

Un dels principis de les polítiques adreçades a dades és que encara que es demani l'obertura per defecte hi ha casos on les dades han de romandre tancades. No es poden posar a disposició de qualsevol persona atesa la seva naturalesa o perquè s'han obtingut de manera confidencial o per temes de seguretat. Tots aquests aspectes han de quedar reflectits en qualsevol pla de gestió de dades que s'elabori. En aquest document s'ha de senyalar tots els procediments que cal seguir quan es treballa amb dades sensibles: aprovacions de comitès d'ètica i/o bioètica, documents de consentiment precisos i indicant clarament quin ús es farà de les dades recollides, com i qui podrà accedir a aquestes dades, i on es dipositaran les dades per vetllar perquè no hi hagi accessos no autoritzats. També cal dir que es pot pensar a compartir les dades de manera agregada o anonimitzant-les però caldrà anar amb compte perquè la desanonimització no sigui senzilla¹⁹.

En els documents per elaborar el pla de gestió de dades del programa H2020 ja s'estableix un llistat d'excepcions: per exemple, quan les dades han de ser explotades comercialment, o quan

hi intervenen aspectes de confidencialitat, seguretat, o quan es treballa dades personals. En tots aquests casos no s'aplica el requeriment de compartir les dades de manera oberta.

Però en els casos on calgui tancar l'accés a les dades, estaria bé obrir l'accés a les metadades. És a dir, que el públic conegués l'existència de les dades i els procediments per poder-hi accedir de manera controlada, sempre que fos possible. Aquest sistema és el que proposa el projecte Privacy Tools de la Universitat de Harvard amb l'etiquetatge del grau de sensibilitat de les dades²⁰. A més d'establir una gradació també ofereixen un qüestionari guiat en línia per tal que els investigadors puguin determinar si les dades amb les quals treballen han d'estar tancades o no. Aquest qüestionari està basat en la legislació dels EUA per tant caldria adaptar-lo a cada país per poder ajustar millor les respostes i incorporar-hi les lleis, les normatives i els reglaments aplicables en cada cas.

5. Conclusió

Hi ha diverses raons per compartir les dades generades o utilitzades al llarg d'una investigació. Però la principal, en el cas de la recerca finançada amb fons públic, és la de facilitar l'accés al públic que l'ha finançada. Aquest retorn és un dels objectius utilitzats per les principals agències de finançament per requerir que les dades es difonguin en obert. A més, l'accés a les dades pot permetre una validació més ràpida dels resultats obtinguts així com la possibilitat de reproduir-los. Però aquest impuls per obrir les dades sempre ha d'anar acompanyat d'un conjunt de mesures per fer possible el tancament de les dades que no es puguin compartir de manera justificada. Caldrà un temps perquè els investigadors s'acostumin aquest nou marc de difusió dels resultats i s'estableixin els procediments per avaluar-lo. Mentrestant ens cal interioritzar que ja no cal explicar perquè volem compartir alguna cosa sinó que hem de justificar perquè la volem tancar: les dades han de ser tan obertes com sigui possible i tan tancades com sigui necessari.

Referències

1. "Budapest Open Access Initiative" [En línia], 14 de febrer de 2002. Disponible a: <http://www.budapestopenaccessinitiative.org/read> [Consulta: 10 de juny de 2017].
2. TENNANT, J.P. *et al.* "The academic, economic and societal impacts of Open Access: an evidence-based review". *F1000Research* 2016, 5:632. Disponible a: [10.12688/f1000research.8460.1](https://doi.org/10.12688/f1000research.8460.1) [Consulta: 10 de juny de 2017].

3. "Ten years on from the Budapest Open Access Initiative: setting the default to open" [En línia], 12 de setembre de 2012, Disponible a: <http://www.budapestopenaccessinitiative.org/boai-10-recommendations> [Consulta: 10 de juny de 2017].
4. "Consultation on 'Science 2.0': Science in Transition" [En línia], 2014, Disponible a: https://ec.europa.eu/research/consultations/science-2.0/consultation_en.htm [Consulta: 10 de juny de 2017].
5. "Amsterdam Call for Action on Open Science" "Amsterdam Call for Action on Open Science" [En línia], 7 d'abril de 2016, Disponible a: <https://english.eu2016.nl/binaries/eu2016-en/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science/amsterdam-call-for-action-on-open-science.pdf> [Consulta: 10 de juny de 2017].
6. THE WELLCOME TRUST, "The Reckoning: An Analysis of Wellcome Trust Open Access Spend 2013-14" [En línia], 3 març 2015, Wellcome Trust Blog, Disponible a: <https://blog.wellcome.ac.uk/2015/03/03/the-reckoning-an-analysis-of-wellcome-trust-open-access-spend-2013-14/> [Consulta: 10 de juny de 2017].
7. PLOS, "Data Availability" [En línia], Disponible a: <http://journals.plos.org/plosone/s/data-availability> [Consulta: 10 de juny de 2017].
8. E-INFRASTRUCTURES AUSTRIA, "Researchers and Their Data. Results of an Austrian Survey - Report 2015". Disponible a: [10.5281/zenodo.34005](https://zenodo.org/record/34005) [Consulta: 10 de juny de 2017]
9. LONGO, D.L.; DRAZEN, J.M.: "Data Sharing", *New England Journal of Medicine*, 2016; vol. 374, p. 276-277. Disponible a: [10.1056/NEJMe1516564](https://doi.org/10.1056/NEJMe1516564) [Consulta: 10 de juny de 2017].
10. KINGSLEY, D.: "We are going OPEN – the Open Research experiment has begun!" [En línia], 8 de febrer de 2017, *Unlocking Research Blog, University of Cambridge Office of Scholarly Communication*. Disponible a: <https://unlockingresearch.blog.lib.cam.ac.uk/?p=1296> [Consulta: 10 de juny de 2017].
11. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS: "Towards better access to scientific information: Boosting the benefits of public investments in research" [En línia], 17 de juliol de 2012, Disponible a: https://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf [Consulta: 10 de juny de 2017].

12. EUROPEAN COMMISSION: "Scientific data: open access to research results will boost Europe's innovation capacity" [En línia], 17 de juliol de 2012, Press Release. Disponible a: http://europa.eu/rapid/press-release_IP-12-790_en.htm?locale=en [Consulta: 10 de juny de 2017].
13. EUROPEAN COMMISSION: "Horizon 2020: Work Programme update supports competitiveness through open science" [En línia], 26 de juliol de 2016, Fact Sheet. Disponible a: http://europa.eu/rapid/press-release_MEMO-16-2604_en.htm?locale=en [Consulta: 10 de juny de 2017].
14. EUROPEAN COMMISSION: "Guidelines on FAIR Data Management in Horizon 2020" [En línia] Versió 3.0 (26 de juliol de 2016). Disponible a: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [Consulta: 10 de juny de 2017].
15. DMP TOOL: "DMP Requirements" [En línia], Disponible a: <https://dmptool.org/guidance> [Consulta: 10 de juny de 2017].
16. DIGITAL CURATION CENTER: "Funders' data plan requirements" [En línia]. Disponible a: <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements> [Consulta: 10 de juny de 2017].
17. WILKINSON, M.D. et al.: "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data*, 2016, vol. 3, Article number: 160018. Disponible a 10.1038/sdata.2016.18 [Consulta: 10 de juny de 2017].
18. CARROLL, M.W.: "Sharing Research Data and Intellectual Property Law: A Primer", *PLoS Biology*, 2015, vol. 13, num. 8, p. e1002235. Disponible a: 10.1371/journal.pbio.1002235.
19. CORTI, L. et al.: "Managing and Sharing Research Data: A Guide to Good Practice", *SAGE*, 2014
20. SWEENEY, L.; CROSAS, M.; BAR-SINAI, M.: "Sharing Sensitive Data with Confidence: The Datatags System" [En línia], *Technology Science*, 2015. Disponible a: <http://techscience.org/a/2015101601/> [Consulta: 10 de juny de 2017].

Fecha de recepció: 3 de junio de 2017

Fecha de aceptació: 28 de junio de 2017