

Les dades de la recerca: de la foscor a la claror

Ignasi Labastida

CRAI-Unitat de Recerca, Universitat de Barcelona
ilabastida@ub.edu

Resum:

Un cop consolidades les polítiques d'accés obert en els resultats de la recerca en publicacions periòdiques, han començat a sorgir accions per obrir altres resultats de recerca com ara les dades obtingudes o emprades. Tradicionalment, aquestes dades només eren accessibles a un grup reduït d'investigadors, però ara es planteja fer-les públiques i oferir-les a qualsevol que les vulgui utilitzar per reproduir la recerca o verificar-ne els resultats. Tanmateix, aquesta obertura de les dades comporta molts dubtes i moltes preguntes per part dels investigadors i les institucions de recerca. No només cal superar les reticències tradicionals, sinó que també cal determinar quines dades s'han de compartir, com s'ha de fer, on s'han de dipositar... En aquest procés, les biblioteques de recerca tenen una nova oportunitat per ajudar a respondre aquestes preguntes i entomar el repte de col·laborar a difondre els resultats de la recerca en obert.

Paraules clau:

dades de recerca, accés obert, biblioteques de suport a la recerca, dades obertes

Los datos de la investigación: de la oscuridad a la claridad

Resumen:

Una vez consolidadas las políticas de acceso abierto en los resultados de la investigación en publicaciones periódicas, han empezado a surgir acciones para abrir otros resultados de la investigación como los datos obtenidos o utilizados. Tradicionalmente, estos datos sólo eran accesibles a un grupo reducido de investigadores, pero ahora se plantea hacerlos públicos y ofrecerlos a cualquiera que los quiera utilizar para reproducir la investigación o verificar los resultados. Sin embargo, esta apertura de los datos comporta muchas dudas y muchas preguntas por parte de los investigadores y las instituciones de investigación. No sólo hay que superar las reticencias tradicionales, sino que también hay que determinar qué datos hay que compartir, cómo hay que hacerlo, dónde hay que depositarlos... En este proceso, las bibliotecas de investigación tienen una nueva oportunidad para ayudar a responder estas preguntas y asumir el reto de colaborar a difundir los resultados de la investigación en abierto.

Palabras clave:

datos de investigación, acceso abierto, bibliotecas de apoyo a la investigación, datos abiertos

Research data: from darkness to clarity

Abstract:

Once open access policies on the publication of the research results have been consolidated, further actions to open more results arisen from research have begun to appear. Among those actions we find policies towards data used or obtained in a research process. Traditionally, those data were only available to a small group of researchers, but now it has been considered to provide them to the public and therefore anyone could use them to reproduce or verify the results. Nevertheless, the opening of data brings many doubts and many questions from researchers and from research institutions. Not only to overcome the traditional reluctance, but also to determine which data to share, how to share them, where to post them ... In this new opening process, research libraries have a new opportunity to help to answer all those questions and take the challenge to collaborate in disseminating research results openly.

Keywords:

data research, open access, research libraries, open data

1. Introducció

Les dades que es fan servir en l'àmbit de la recerca han estat tradicionalment accessibles només a un grup reduït de persones. Els investigadors han estat molt gelosos de les seves dades i normalment les han tingut tancades [1]. Les raons per no compartir-les han estat variades: des de la por a perdre la primícia o que siguin mal interpretades fins a la necessitat de dedicar temps a polir-les o preparar-les per fer-les públiques¹. Tanmateix, els darrers anys aquesta tendència està canviant lentament. Els investigadors han vist com durant la primera dècada del segle XXI s'ha estès el moviment de l'accés obert que aposta per fer accessible de manera gratuïta la publicació dels resultats de la recerca i permetre'n la difusió sense barreres tecnològiques ni legals². Les institucions públiques i privades que financen la recerca han establert una sèrie de polítiques³ imposant que aquests resultats es posin a l'abast de tothom en un temps no superior a un any després de ser publicats en revistes de referència. El pas següent ha estat implementar aquestes polítiques a altres resultats de la recerca: les dades. Per tant, aquest canvi de tendència pel que fa a la difusió de les dades no ha vingut principalment per part dels investigadors sinó per qui en finança la recerca o per part de les mateixes publicacions. El nombre de revistes que demanen als investigadors la difusió de les dades obtingudes o emprades en una recerca abans de decidir si aproven o no la publicació d'un resultat ha augmentat els darrers anys [2]. Aquests canvis comporten l'aparició de dubtes i preguntes sobre quines dades cal oferir, on s'han d'ubicar, com s'han de citar o si hi haurà algun reconeixement específic per als investigadors que ho facin com en el cas de les publicacions, o quins beneficis comporta. Sobre aquest darrer punt, els beneficis, s'han escrit diversos estudis. N'hi ha alguns que apunten que compartir les dades augmenta el nombre de



©istockphoto/sonyae

citacions de les publicacions relacionades [3] i d'altres que apunten també als beneficis de la institució per aconseguir noves col·laboracions o nous fons [4].

En aquest article revisarem els aspectes a tenir en compte a l'hora de difondre les dades de recerca de manera oberta. Veurem què s'està fent en l'àmbit internacional i en l'estatal. I finalment, analitzarem el paper que hi pot jugar una biblioteca o un centre de recursos d'una institució de recerca. Comencem per veure què entenem per dada de recerca.

1.1 Les dades de la recerca

No és fàcil parlar de dades de recerca perquè cadascú pot tenir la seva interpretació del que és una dada segons la disciplina. Per a un meteoròleg una dada pot ser la tempe-

1. Hi ha un document dinàmic a la xarxa que intenta donar resposta a alguna d'aquestes raons: Christopher Gutteridge, Alexander Dutton, «Concerns about opening up data, and responses which have proved effective» [En línia]. <https://docs.google.com/document/d/1nDtHpnIDTY_G32EMJni-XaOGBufjHCck4VC9WGO7jK4/edit#> [Consulta: 1/7/2013].
2. «Iniciativa de Budapest per a l'accés obert» [En línia], febrer de 2002. <<http://www.budapestopenaccessinitiative.org/translations/catalan-translation>> [Consulta: 1/7/2013].
3. Es poden consultar aquestes polítiques a www.sherpa.ac.uk/juliet [Consulta: 1/7/2013].

ratura en un determinat lloc i en un determinat moment, però per a un metge una dada pot ser una imatge obtinguda en un aparell de ressonància magnètica. De fet, si consultem el *Gran Diccionari de la Llengua Catalana* ens trobem com a primera definició de dada: «Qualsevol informació que facilita la formació d'una idea, d'un judici, d'una conclusió»⁴.

Aquesta varietat de significats té la seva rellevància per tal d'establir plans de gestió de dades o per determinar l'estat legal d'una dada, com veurem a continuació.

Podem classificar les dades d'acord amb la seva procedència, i llavors parlarem de dades primàries i dades secundàries. Les dades primàries són aquelles dades que obtenim per l'observació o l'experimentació; en canvi, les dades secundàries són aquelles que han estat tractades i que poden ser fruit de la combinació de diversos conjunts de dades primàries.

El que tractarem en aquest article serà bàsicament les dades primàries, que podem separar en dades observacionals fenomenològiques, experimentals o computacionals, segons quin sigui el seu origen [5].

1.2 Estat legal de les dades

Com ja s'ha dit anteriorment, l'estat legal de les dades dependrà del que entenguem com a dada. En el cas que considerem com a dada un document, una imatge o un registre sonor, haurem de tenir en compte que poden gaudir de protecció per la propietat intel·lectual i, per tant, caldrà tenir en compte qui n'és el titular i quins usos poden estar permesos. El titular o beneficiari dels drets de propietat intel·lectual pot ser l'investigador, el centre de recerca o

Quan la dada no tingui la consideració per ser protegida per la propietat intel·lectual, com és el cas de moltes dades primàries, el fet de fer-ne una compilació o una base de dades sí que genera uns drets.

el finançador de la recerca, segons estigui estipulat per llei, per conveni o acord, o segons l'establert en l'ajut per realitzar la recerca. Cal establir clarament qui és o serà el titular d'aquestes dades des de bon principi per poder determinar qui decidirà com es difondran⁵.

Quan la dada no tingui la consideració per ser protegida per la propietat intel·lectual, com és el cas de moltes dades primàries, el fet de fer-ne una compilació o una base de dades sí que genera uns drets. Aquesta consideració pot variar segons les lleis vigents a cada país [6]. A la Unió Europea, una base de dades pot ser protegida com una obra artística, científica o literària sempre que la selecció dels continguts o la seva disposició sigui original. I en el cas que no ho sigui, el sol fet d'invertir temps, diners o energia en la creació de la base dades també atorga un dret al seu creador. Aquest dret és l'anomenat dret *sui generis*⁶, que permet al creador de la base dades prohi-

- «Dada» a Grup Enciclopèdia Catalana, *Gran Diccionari de la Llengua Catalana* [En línia]. Barcelona: 1a edició, 1998. <<http://www.diccionari.cat/lexicx.jsp?GECART=0157101>> [Consulta: 1/7/2013].
- Un exemple de política de dades el trobem al projecte IceCube (icecube.wisc.edu/science/data) [Consulta: 1/7/2013].
- Introduït per la *Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos* [En línia]. <http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexapi!prod!CELEXnumdoc&lg=en&numdoc=31996L0009&model=guichett> [Consulta: 10/8/2013].

bir-ne l'extracció o la reutilització d'una part substancial o la totalitat dels continguts. Per tant, com en el cas anterior, on els continguts sí que eren protegibles, també caldrà establir qui té la titularitat d'aquests drets.

2. Aspectes a tenir en compte a l'hora de difondre i compartir les dades

A l'hora de compartir les dades de la recerca no només cal tenir en compte què entenem per dades i quin és el seu estat legal, sinó que també cal avaluar altres aspectes com són: la selecció de les dades, la descripció de les dades, com indicar-ne l'obertura, la seva conservació i preservació i, finalment, la citació de les dades

2.1 Selecció de les dades

El nombre de dades que tenim al nostre abast creix de manera desmesurada dia a dia. Ens trobem davant de l'anomenat «diluvi» o «allau de dades» [7] i, per tant, és lògic plantejar-se quines dades cal compartir. Cal fer una preselecció de les dades amb les quals s'ha treballat al llarg d'una investigació? O s'haurien d'exposar totes les dades per tal que qualsevol persona pogués verificar-les i reutilitzar-les? No hi ha una resposta única per a tots els casos i totes les disciplines. A més, la determinació de quines dades cal difondre pot estar fixada per una política de la institució, agència o administració que finança la recerca. Tanmateix, es poden establir algunes pautes per discernir quines dades cal difondre [8]. Entre els criteris que es poden seguir hi ha la importància de les dades i del context on han estat obtingudes o observades; el grau de reutilització de les dades, és a dir, si tenen un format que en permetrà la reutilització posterior; la relació amb altres dades importants, o el fet que les dades són úniques i

difícilment es podran tornar a recollir. També hi ha un altre punt clau, el de la dimensió de les dades i el cost que pot tenir conservar-les. Per aquesta raó, és important també determinar si les dades seleccionades mantindran la seva validesa i importància al llarg del temps.

És important seleccionar les dades a priori, és a dir, determinar al principi de la recerca quines dades seran les que es difondran i es compartiran. Aquest punt serà l'inici d'un pla de gestió de dades que és necessari elaborar en qualsevol recerca si al final volem tenir-les en obert. Tampoc no podem oblidar que en alguns àmbits s'utilitzen dades personals o confidencials i, per tant, caldrà tenir previst un protocol o procediment per anonimitzar les dades o excloure-les de la difusió.

2.2 Descripció de les dades

Un cop determinades les dades que es compartiran, caldrà descriure-les perquè posteriorment es puguin citar, identificar o conèixer com han estat obtingudes. Hi ha diferents models per descriure les dades, però potser caldria veure aquells que han estat creats especialment per a les dades com, per exemple, el model proposat per la Data Documentation Initiative (DDI)⁷ per al camp de les ciències socials.

Com a informació bàsica caldria donar un títol, el nom de les persones que han obtingut les dades, un identificador, un nombre de matèries, la situació legal, les dates més rellevants, com ara la d'obtenció o processament de les dades, i el format. També caldrà indicar, si és possible, qui ha finançat la recerca, on s'han obtingut les dades o quina metodologia s'ha seguit. El nombre de metadades que cal per descriure un conjunt de dades vindrà donat molts cops pel lloc on s'allotjaran i que marcarà quines són obligatòries i quines són recomanables, a més d'oferir estàndards ja establerts.

7. Data Documentation Initiative (www.ddialliance.org) [Consulta: 1/7/2013].

2.3 L'obertura de les dades

Per obrir les dades caldrà indicar-ho, ja que el marc legal actual estableix que no dir res equival a reservar-se tots els drets⁸ i, per tant, haver de demanar permís per utilitzar l'obra o creació, llevat dels casos previstos en la llei. Aquests casos són els límits o excepcions dels drets de propietat intel·lectual als quals podem recórrer en determinades condicions en el marc de la recerca, però són molt detallats.

Així doncs, si volem compartir les dades caldrà indicar-ho mitjançant algun tipus d'avís legal. Possiblement el mitjà més estès actualment són les llicències per a continguts oberts. Aquestes llicències són textos legals mitjançant els quals els autors o titulars d'una obra o creació, de la qual ostenten els drets, només se'n reserven alguns i permeten, a priori, determinats usos amb algunes condicions. De llicències d'aquest tipus n'hi ha moltes, però possiblement el sistema més estès actualment és el de Creative Commons. Quan les dades que vulguem obrir puguin ser protegides per propietat intel·lectual com una obra artística, científica o literària podem fer servir alguna de les sis llicències que ofereix Creative Commons. No obstant això, si volem complir estrictament la definició d'accés obert caldria utilitzar la llicència de Reconeixement (CC BY)⁹, que permet qualsevol tipus d'ús sense restriccions sempre que se citi l'autoria i sigui visible la indicació de la llicència.

En el cas que l'objecte protegible per propietat intel·lectual no siguin les dades sinó la base de dades, les sis llicèn-

cies estàndard de Creative Commons només servien en el cas que la base de dades fos original, perquè en el text legal hi ha una renúncia explícita a qualsevol *dret sui generis*¹⁰. Aquest fet implica que no té sentit posar una llicència d'aquest tipus en moltes de les bases de dades de recerca que es generen, ja que no tenen el grau d'originalitat suficient per ser protegides més enllà d'aquest *dret sui generis* aplicable. Així doncs, decidir quina llicència s'adapta millor a una base de dades no és fàcil i, de fet, a Creative Commons es va obrir una discussió sobre això¹¹ que va desembocar en l'elaboració d'una nova llicència, CC Zero (CC0)¹², i a l'establiment d'un protocol per implementar l'accés obert a dades¹³. La llicència CC0 és, de fet, una renúncia a exercir qualsevol dret per part del titular de la propietat intel·lectual. Aquesta llicència va més enllà de la llicència de Reconeixement (CC BY) perquè ni tan sols requereix el reconeixement d'autoria, segons la legislació aplicable, sinó que demana que se segueixin les normes de citació de la comunitat científica. Aquesta llicència o renúncia és la que actualment es fa servir per les metadades del projecte Europeana¹⁴.

La fundació Open Knowledge també ofereix unes eines legals per obrir les dades mitjançant el seu projecte Open Data Commons¹⁵. Hi ha tres llicències dissenyades per a bases de dades, i que són compatibles amb el protocol de Creative Commons esmentat anteriorment. Al contrari que les sis llicències genèriques de Creative Commons, aquestes llicències han estat creades per tenir en compte les peculiaritats de les bases de dades, com per exemple el *dret sui generis* aplicable a la Unió Europea. Una de les tres llicències és molt similar a la llicència CC0, mentre

8. L'article 1 de l'actual llei de propietat intel·lectual vigent a l'Estat espanyol estableix que no cal cap tràmit per gaudir dels drets: «La propietat intel·lectual d'una obra literària, artística o científica correspon a l'autor per l'únic fet de la creació». Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el Texto Refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia, *Boletín oficial del Estado*, 22 de abril de 1996, n.º. 97, p. 14369-14396.
9. OASPA, «Why CC BY?» [En línia]. <<http://oaspa.org/why-cc-by>> [Consulta: 1/7/2013].
10. Cal tenir en compte que als EUA el *dret sui generis* per a bases de dades no està previst.
11. Creative Commons, «Data» [En línia]. <<http://wiki.creativecommons.org/Data>> [Consulta: 1/7/2013].
12. Creative Commons, «About CC0 - "No Rights Reserved"» [En línia]. <<http://creativecommons.org/about/cc0>> [Consulta: 1/7/2013].
13. Science Commons, «Protocol for Implementing Open Access Data» [En línia]. <<http://sciencecommons.org/projects/publishing/open-access-data-protocol>> [Consulta: 1/7/2013].
14. Europeana, «Data Exchange Agreement» [En línia], juliol 2012. <<http://pro.europeana.eu/web/guest/data-exchange-agreement>> [Consulta: 10/8/2013].
15. Open Data Commons (opendatacommons.org) [Consulta: 1/7/2013].



©istockphoto/tommytucker

que les altres dues inclouen el reconeixement d'autoria, en un cas, i la condició *copyleft* en l'altre, que requereix mantenir la llicència en qualsevol adaptació que es faci de la base de dades, seguint la filosofia del programari lliure.

Un altre referent per entendre com obrir les dades són els anomenats *Principis de Panton*¹⁶, un document elaborat l'any 2010 al Regne Unit per un grup d'experts en dades obertes, i que estableix els principis per considerar unes dades com a obertes.

2.4 Conservació i preservació

Quan les dades ja han estat seleccionades, identificades i obertes cal col·locar-les en un lloc on estiguin disponibles i localitzables de manera perdurable. Generalment no és convenient deixar-les en pàgines personals, sinó que cal ubicar-les en repositoris dedicats a aquest efecte o que al-

menys permetin la interconnexió amb d'altres mitjançant protocols estàndard.

A l'hora de decidir en quin repositori cal dipositar les dades reapareix el debat sobre el repositoris locals i els globals. Cal crear repositoris locals de dades? O bé cal treballar a nivells superiors? Una primera opció seria dipositar les dades en el repositori institucional on s'ha dut a terme la recerca¹⁷. Tanmateix, el cost de preservar aquestes dades i el nombre creixent de les dades pot fer inviable aquesta primera solució i, per tant, s'hauria d'optar per opcions consorciades a un nivell superior [9]. En aquest segon cas ens trobem amb infraestructures temàtiques, és a dir, repositoris per a àmbits o disciplines de recerca que permeten una adaptació de les metadades a les necessitats de cada camp. Una altra possibilitat és crear repositoris nacionals o continentals que estiguin connectats amb altres de similars i que permetin una visió més global amb subdivisions temàtiques especialitzades.

16. El text original en anglès i la traducció al català es troben disponibles a Panton Principles, «Principis per obrir les dades científiques» [En línia]. <<http://pantonprinciples.org>> [Consulta: 1/7/2013].

17. Segons el registre de repositoris d'accés obert (*roar.eprints.org*), a principis de juliol de 2013 hi havia uns 3.400 repositoris al món.

Algunes d'aquestes solucions ja s'han posat en marxa a diferents nivells i actualment ja es poden dipositar les dades en repositoris d'accés públic. Per ajudar a prendre una decisió es pot consultar un parell de directoris de repositoris de dades¹⁸. Cal dir que el nombre de revistes que requereixen que les dades estiguin dipositades en algun repositori públic i, per tant, siguin accessibles va en augment i moltes aconsellen consultar aquests directoris per trobar el millor lloc on hostatjar-les¹⁹.

2.5 Citació

Finalment, cal veure com se citen les dades. Generalment, aquest tema està vinculat amb la ubicació de les dades. Quan les dades es troben en un repositori normalment hi ha la possibilitat d'utilitzar el sistema Handle²⁰, que és el més estès en aquest tipus de plataformes. No obstant això, molts repositoris que ofereixen un servei gratuït d'hostatjament de dades, de vegades vinculat a la publicació en revistes, disposen del sistema d'identificació DOI, el mateix que s'utilitza en la majoria de publicacions periòdiques internacionals. Entre aquests repositoris que utilitzen el DOI hi ha Dryad²¹ o Figshare²². Aquests identificadors estan pensats per garantir un accés permanent a l'objecte associat. A més, hi ha altres mètodes per identificar unívocament un conjunt de dades com, per exemple, l'empremta numèrica universal o UNF (Universal Numerical Fingerprint) [10]. Aquest identificador es crea mitjançant un algoritme per preservar la integritat del conjunt de dades i poder verificar que no ha canviat encara que hagi variat el lloc on ha estat emmagatzemat o allotjat.

I un cop les dades són citables es poden construir mètriques per avaluar-ne les citacions, com s'ha fet durant

anys amb les publicacions [11]. Hi ha diversos models proposats, però encara no hi ha un estàndard adoptat. En el que tothom coincideix és que cal buscar un sistema per recompensar i incentivar els investigadors més enllà de la bona voluntat de compartir les dades [12].

3. Situació de l'obertura de dades de recerca en el context internacional

En l'àmbit internacional hi ha tres zones on s'està treballant de manera més activa per obrir les dades de recerca. Per una banda hi ha els EUA, on diferents agències i instituts que financen la recerca tenen polítiques de gestió de les dades de recerca [13] en què s'inclouen els períodes durant els quals els investigadors poden gaudir de l'exclusivitat d'accés a les dades i els terminis per fer públiques les dades primàries i secundàries. Cal tenir en compte que la llei de propietat intel·lectual dels EUA, Copyright Act, estableix que el que crea el Govern federal passa automàticament al domini públic, és a dir, no té cap mena de protecció de drets²³. Aquesta situació afecta moltes d'aquestes agències o instituts nacionals. En moltes biblioteques i centres de recursos es poden trobar serveis especialitzats adreçats a donar suport als investigadors en la gestió de dades²⁴.

Un segon referent és Austràlia. De fet, aquest país va ser un dels primers a adoptar polítiques per obrir les dades públiques i, per tant, és lògic que aquestes polítiques també s'apliquen en les dades de recerca. El Servei Nacional de Dades Australià ha establert un repositori de dades nacio-

18. El Registre de Repositoris de Dades de Recerca o re3data.org (www.re3data.org) i el Databib (databib.org) [Consulta: 1/7/2013].

19. Per exemple, la política editorial de SpringerPlus (www.springerplus.com/about#editorialpolicies) [Consulta: 1/7/2013].

20. Handle (www.handle.net) [Consulta: 1/7/2013].

21. Dryad és un repositori genèric per hostatjar dades relacionades amb publicacions científiques (datadryad.org) [Consulta: 10/8/2013].

22. Figshare s'ofereix com un espai alternatiu on publicar dades i altres resultats de la recerca al que ofereix l'actual sistema de comunicació científica (figshare.com) [Consulta: 10/8/2013].

23. Secció 105 de la *Copyright Act* [En línia], <<http://www.law.cornell.edu/uscode/text/17/105>> [Consulta: 1/7/2013].

24. Un exemple és el del MIT i el seu servei «Manage your data» (libraries.mit.edu/guides/subjects/data-management/index.html), que ofereix guies, seminaris, atenció personalitzada a grups i investigadors [Consulta: 1/7/2013].

nal, Research Data Australia²⁵, que el juliol de 2013 ja disposava de més de 56.000 conjunts de dades de recerca. Les dades allotjades estan etiquetades amb metadades que inclouen les condicions d'accés i la llicència d'ús, si existeix.

Finalment, el tercer focus és la Unió Europea, on s'està establint una política conjunta de cara al nou programa marc d'ajuts per a la recerca anomenat *Horizon 2020*. En aquest programa, que s'iniciarà l'any 2014, es preveu una acció pilot sobre dades de recerca que requerirà que en determinats àmbits s'hagin de fer públiques les dades amb les quals s'ha treballat en el marc d'un projecte finançat per aquest programa²⁶. Si mirem els desenvolupaments que s'han fet a Europa a nivell estatal possiblement destaca el Regne Unit, on el Govern ha apostat clarament per l'accés obert després de la publicació de l'informe *Finch*²⁷. Com en el cas australià, cal destacar que el Regne Unit és un dels estats més actius en l'obertura de les dades públiques després de la campanya iniciada el març de 2006 reclamant l'alliberament d'aquestes dades²⁸. Pel que fa a les accions respecte les dades de recerca, cal destacar l'informe elaborat per la Royal Society [14] on es recullen una sèrie de recomanacions adreçades a investigadors, centres de recerca, societats científiques, editors, governs i agències de finançament per tal de facilitar la ubicació i la difusió en obert de les dades recollides en qualsevol recerca. Com en el cas dels EUA, la majoria d'universitats disposen de serveis de suport a la gestió d'aquestes dades²⁹.

No voldria acabar aquest apartat sense esmentar l'Aliança per a les Dades de Recerca (Research Data Alliance)³⁰, una

organització internacional impulsada pels organismes de finançament de la recerca a Austràlia, els EUA i la Unió Europea. L'objectiu d'aquesta organització és facilitar i accelerar l'intercanvi de dades de recerca mitjançant la implementació de la infraestructura tecnològica i les connexions socials necessàries. Aquesta aliança s'organitza mitjançant grups de treball i grups d'interès oberts a la participació de qualsevol persona interessada i en trobades semestrals plenàries.

4. Situació de les dades a l'Estat espanyol

A finals de l'any passat, la FECYT (Fundación Española para la Ciencia y la Tecnología) va publicar un informe del grup de treball sobre dipòsit i gestió de dades en accés obert [14]. Possiblement aquest document és un bon inici per establir una política de gestió de dades en qualsevol institució de recerca. S'hi apunten algunes iniciatives estatals, com ara el projecte ODISEA³¹. Aquest projecte té com a objectiu crear un inventari internacional dels repositoris que admeten el dipòsit de dades de recerca. A més d'aquest projecte, s'esmenten les publicacions d'autors estatals sobre el tema de les dades de recerca, es fa un repàs de les sessions, jornades o conferències organitzades per difondre l'estat de les dades de recerca i conscienciar els investigadors, i finalment se citen projectes de recerca relacionats amb les dades.

25. Research Data Australia (researchdata.andis.org.au) [Consulta: 1/7/2013].

26. «Towards better access to scientific information: Boosting the benefits of public investments in research» [En línia]. Communication from the European Commission, 17 de juliol 2012. <http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf> [Consulta: 1/7/2013].

27. Resposta del govern britànic a la publicació de l'informe: Letter to Dame Janet Finch on the Government Response to the Finch Group Report: «Accessibility, sustainability, excellence: how to expand access to research publications» [En línia]. <<http://www.bis.gov.uk/assets/biscore/science/docs/l/12-975-letter-government-response-to-finch-report-research-publications.pdf>> [Consulta: 1/7/2013].

28. Free Our Data: Make taxpayers' data available to them (www.freeourdata.org.uk) [Consulta: 1/7/2013].

29. Un exemple d'això és el servei ofert per la Universitat de Cambridge: Support for Managing Research Data (www.lib.cam.ac.uk/dataman) [Consulta: 1/7/2013].

30. Research Data Alliance (rd-alliance.org) [Consulta: 1/7/2013].

31. ODISEA, International Registry on Research Data (odisea.ciepi.org) [Consulta: 1/7/2013].



©istockphoto/Steven Frame

Tanmateix, no hi ha cap política adoptada per una institució de recerca o una administració pública que requereixi cap tipus de dipòsit de dades. Tampoc no hi ha cap iniciativa per construir un repositori estatal. En l'informe tampoc no s'esmenten serveis de suport a la recerca com els que hi ha als EUA o al Regne Unit. Això no vol dir que no hi hagi interès a crear-los, sinó que possiblement encara no estan prou consolidats per posar-los com a exemple.

5. El paper d'una biblioteca o un centre de recursos

Des de sempre les biblioteques o els centres de recursos han ofert serveis de suport als investigadors, des de l'accés a les publicacions fins a l'ajuda per fer cerques bibliogràfiques. Aquests centres sempre han estat un dels principals punts d'informació als quals els investigadors s'han adreçat. Els darrers anys, un dels temes més tractats ha estat l'accés obert i tot el que comporta: repositoris, revistes, ajudes, aspectes legals... La majoria d'institucions de

Des de sempre les biblioteques o els centres de recursos han ofert serveis de suport als investigadors, des de l'accés a les publicacions fins a l'ajuda per fer cerques bibliogràfiques.

recerca han construït unitats de suport a la recerca constituïdes principalment per bibliotecaris i documentalistes, sense oblidar altres perfils. Cal no ignorar tampoc les connexions necessàries amb altres unitats de la institució com ara les oficines de gestió de la recerca o de transferència, els centres d'informàtica o els serveis de publicacions. Sense una estratègia global i coordinada, la institució no podrà oferir un suport eficient als investigadors. Ara, el repte que es presenta és col·laborar i donar suport a la gestió i la difusió de les dades de recerca [15]. En aquest camp, les biblioteques o els centres de recursos poden oferir tota la seva expertesa en la gestió de la informació i l'experiència adquirida en la darrera dècada amb el moviment de l'accés obert. Un bon punt de partida per establir un suport és seguir les recomanacions de LIBER [16]. En aquestes recomanacions s'inclou la possibilitat d'oferir suport específic per a la gestió de les dades amb ajudes centrades en els temes legals o en la descripció dels conjunts de dades mitjançant metadades, la creació de llocs de treball o perfils específics dedicats a la gestió i conservació de les dades, la participació en les polítiques institucionals, l'oferiment de col·laboracions amb investigadors i grups de recerca, i l'oferiment d'espais on allotjar

les dades, com per exemple els repositoris institucionals, o l'aliança amb altres serveis o institucions per construir infraestructures comunes.

Un dels serveis més citats en el suport que ofereixen les biblioteques o centres de recursos anglosaxons és ajudar en l'elaboració dels plans de gestió de dades. Cada cop més, qui finança la recerca demana l'elaboració d'aquests plans i previsiblement serà així en la nova convocatòria d'ajuts europeus. Un pla de gestió de dades hauria d'incloure la tipologia de dades, les mostres i els materials que s'utilitzaran i es produiran al llarg de la recerca, així com el programari o altres eines que es faran servir. Cal incloure-hi els estàndards que se seguiran pel que fa a la descripció de les dades mitjançant metadades i els formats utilitzats. S'hi han d'afegir les polítiques d'accés i reutilització sense oblidar l'estat de la propietat intel·lectual o la protecció de les dades per temes de seguretat, confidencialitat o privacitat. I finalment, cal establir un pla d'arxivament i preservació de les dades a llarg termini.

A més, des de les biblioteques o centres de recursos es poden realitzar campanyes de sensibilització i informació per tal d'arribar als investigadors de la institució.

6. Conclusions

Com a conclusió m'agradaria tornar a remarcar que l'obertura de les dades de recerca ofereix moltes oportunitats per a investigadors, institucions de recerca, centres de recursos, emprenedors i ciutadans. Per una banda l'investigador pot aconseguir una major visibilitat dels resultats de la seva recerca i, per tant, augmentar el seu reconeixement i les seves citacions. La institució on treballa pot aconseguir més prestigi i establir noves col·laboracions amb centres de recerca d'arreu del món. Per a la biblioteca o el centre de recursos és una oportunitat per adaptar-se als canvis, establir nous perfils, entomar nous reptes... Els emprenedors poden desenvolupar nous projectes a partir de les dades obertes. I, finalment, al ciutadà li oferim la

possibilitat d'acostar-lo a la recerca que es desenvolupa al seu entorn i que en la majoria de casos ell mateix està finançant. Ha arribat el moment que les dades de la recerca vegin la llum!

Referències

- [1] TENOPIR, C.; ALLARD, S.; DOUGLASS, K.; AYDINOGLU, A. U.; WU, L.; *et al.* (2011) «Data Sharing by Scientists: Practices and Perceptions». PLoS ONE 6(6): e21101. DOI:10.1371/journal.pone.0021101.
- [2] STODDEN, V.; GUO, P.; MA, Z. (2013) «Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals». PLoS ONE 8(6): e67111. DOI:10.1371/journal.pone.0067111.
- [3] PIWOWAR, H. A.; DAY, R. S.; FRIDSMA, D. B. (2007) «Sharing Detailed Research Data Is Associated with Increased Citation Rate». PLoS ONE 2(3): e308. DOI:10.1371/journal.pone.0000308
- [4] WOLSKI, M.; RICHARDSON, J.; REBOLLO, R.; «Shared benefits from exposing research data», 32nd Annual IATUL Conference Proceedings, 2011.
- [5] TORRES-SALINAS, D.; ROBINSON-GARCÍA, N.; CABEZAS-CLAVIJO, A. «Compartir los datos de investigación: introducción al data sharing». *El profesional de la información*, 2012, marzo-abril, vol. 21, núm. 2, p. 173-184. DOI:10.3145/epi.2012.mar08.
- [6] DE COCK BUNING, M.; VAN DINTHER, B.; JEPPESEN DE BOER, C. G.; Ringnald, A. «The legal status of research data in the Knowledge Exchange partner countries» [En línia], Centre for Intellectual Property Law (CIER), Molengraaff Institute for Private Law, Utrecht University, The Netherlands, octubre de 2011. <<http://www.knowledge-exchange.info/Default.aspx?ID=461>> [Consulta: 1/7/2013].

- [7] HEY, Tony; TREFETHEN, Anne. «The Data Deluge: an e-Science Perspective» dins *Grid Computing: Making the Global Infrastructure a Reality*, Fran Berman, Geoffrey C. Fox, and Anthony J. Hey (eds.), 2003, p. 808-822.
- [8] TJALSMA, Heiko; ROMBOUTS, Jeroen. *Selection of Research Data. Guidelines for appraising and selecting research data. A report by DANS and 3TU.Datacentrum* [En línia], Surf Foundation, 2010. <http://www.surf.nl/nl/themas/openonderzoek/cris/Documents/SURFshare_Collectioneren_Selection%20of%20Research%20Data_DANS_3TU_DEFtt.pdf> [Consulta: 1/7/2013].
- [9] BEAGRIE, N.; CHRUSZCZ, J.; LAVOIE, B. *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities* [En línia]. JISC, 2008. <<http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>> [Consulta: 1/7/2013].
- [10] ALTMAN, M., KING, G. A. «Proposed Standard for the Scholarly Citation of Quantitative Data» [En línia], *D-Lib Magazine*, 2007, vol.13, núm. 3/4. <<http://www.dlib.org/dlib/march07/altman/03altman.html>> [Consulta: 1/7/2013].
- [11] COSTAS, R.; MEIJER, I.; ZAHEDI, Z.; WOUTERS, P. *The Value of Research Data - Metrics for datasets from a cultural and technical point of view. A Knowledge Exchange Report* [En línia]. Leiden: 2013. <<http://www.knowledge-exchange.info/datametrics>> [Consulta: 1/7/2013].
- [12] NIU, Jinfang. «Reward and Punishment Mechanism for Research Data Sharing», *IASSIST Quarterly*, No. Winter, 2006.
- [13] DIETRICH, D.; ADAMUS, T.; MINER, A.; STEINHART, G. «De-Mystifying the Data Management Requirements of Research Funders», *Issues in Science and Technology Librarianship*, 70, 2012. DOI:10.5062/F44M92G2.
- [14] The Royal Society. *Science as an open enterprise*. London: The Royal Society Science Policy Centre report, 02/12, June 2012. ISBN: 978-0-85403-962-3.
- [15] Grupo de Trabajo de Depósito y Gestión de datos en Acceso Abierto del proyecto RECOLECTA. *La conservación y reutilización de los datos científicos en España. Informe del grupo de trabajo de buenas prácticas* [En línia]. Madrid: Fundación Española para la Ciencia y la Tecnología (FECYT), (2012). <<http://www.fecyt.es>> [Consulta: 1/7/2013].
- [16] MARTÍNEZ-URIBE, L.; MACDONALD, S. «Un nuevo cometido para los bibliotecarios académicos: data curation» [En línia]. *El profesional de la información*, 2008, vol. 17, núm. 3, p. 273-280. <<http://www.elprofesionaldelainformacion.com/contenidos/2008/mayo/03.pdf>> [Consulta: 1/7/2013].
- [17] CHRISTENSEN-DALSGAARD, Birte; *et al.* «Ten recommendations for libraries to get started with research data management» [En línia]. Final report of the LIBER working group on E-Science / Research Data Management, juliol 2012. <<http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf>> [Consulta: 1/7/2013].