

Tendencias internacionales en el acceso libre a la documentación científica digital: OAI-Open Archives Initiative

M^a Fernanda Peset Mancebo
(DCADHA-UPV)
Alfredo Ortín Pérez
(Biblioteca de Humanidades Joan Reglà-UV)

Introducción

En este homenaje a Mariano Peset queremos abordar un tema que compromete la labor de los académicos a escala mundial: el acceso a la información científica. Tres son los agentes que intervienen en el ciclo de la comunicación científica: editores, autores y bibliotecas; que en su mayoría en España, exceptuando los institutos de investigación, se suelen ubicar en el ámbito universitario. Este predominio de las instituciones académicas de educación superior condicionará no sólo las costumbres y los intereses de los creadores de información científica, sino también aspectos relacionados con su gestión y difusión a través de bibliotecas. Esta perspectiva técnica es la que se aborda en este artículo. Las bibliotecas universitarias tienen unas peculiaridades claras respecto a otras en cuestión de gestión de derechos de propiedad intelectual y bibliotecas digitales. Y como hemos comentado, debido a la vertiente investigadora de parte de sus usuarios, cuentan con la perspectiva de los productores o creadores de información científico-técnica. Así, a lo largo de estas páginas se tratará la armonización entre las tendencias en la difusión de la información y la propiedad intelectual de los escritos de los autores científicos.

Las vías tradicionales de difusión de la ciencia

Comencemos sintetizando con cierta profundidad la situación de los autores y la comunicación científica. Las vías tradicionales de difusión de la ciencia han sido desde hace siglos las publicaciones impresas de revistas, actas de congresos o monografías, que eran adquiridas a las editoriales públicas y privadas por los propios autores o sus instituciones -depositándose en último término en las bibliotecas-. Este sistema aseguraba tanto una primera difusión a través de publicaciones que, mediante un sistema de evaluación por pares articulado por las editoriales, reconocían la calidad de los trabajos publicados; la conservación de los títulos de las revistas y los documentos adquiridos por las bibliotecas para dar servicio a sus usuarios; como, por último, una segunda difusión de los trabajos al referenciarse en bases de datos bibliográficas científicas.

Este complejo sistema –autores, editores, bibliotecas- sin embargo, lejos de estar en equilibrio, sufre constantes tensiones en función de los intereses de cada uno de los polos. Autores y bibliotecas suelen además pertenecer al ámbito de la administración pública, en el que, si bien no necesitan buscar constantemente financiación –al menos las universitarias de nuestro país-, sí han de justificar su existencia con resultados cuantificables y con la utilidad de

sus servicios. Por el contrario, el mundo de la edición es mixto, tanto en el caso de revistas como de bases de datos. En ciertas disciplinas vinculadas al sistema de patentes –medicina, ingenierías, farmacia, informática...- los proveedores de esta información son empresas privadas, cada vez más concentradas, además, a escala internacional -Elsevier, Springer-Kluwer, Thompson...-. Mientras que en otras disciplinas, en las que no se genera negocio -humanidades, ciencias sociales...- o en las ciencias básicas no aplicadas –como matemática-, sólo las revistas de mayor impacto están en manos de editores privados. El resto sería editado por sociedades científicas, universidades, gobiernos... La existencia de una lógica mercantil en este escenario donde autores y bibliotecas son de administración pública, como vimos, ha provocado el alza progresiva de los precios de las revistas y la incapacidad de las bibliotecas de asumir el crecimiento de títulos cada vez más onerosos. Puede decirse que las organizaciones que sustentan el trabajo de autores y las bibliotecas que les dan servicio, se ven perjudicados por la introducción de intereses mercantiles en un sistema predominantemente sin ánimo de lucro.

Así, se ha hecho patente una paradoja en la producción y consumo de la información científica, a diferencia de la comercial. El autor científico persigue la máxima visibilidad de su trabajo, lo que repercutirá en un reconocimiento entre la comunidad científica. Y éste es su máximo interés, pues, al menos en nuestro país, los autores científicos no confían en los réditos de sus publicaciones para subsistir, sino que se encuentran adscritos a alguna organización que les paga por su trabajo (universidades, institutos de investigación o empresas). Para García Testal y Barrueco Cruz¹, “Mientras el autor literario obtendrá, además del reconocimiento de la sociedad, una parte proporcional de los beneficios que origine la venta de sus obras, el autor científico técnico no sólo no recibe una remuneración económica directa por su trabajo sino que además, en ocasiones, debe costear total o parcialmente la publicación del mismo. El autor sólo obtiene un beneficio intelectual o moral: el reconocimiento de la comunidad académica. Un reconocimiento que, no obstante, se traducirá en algún tipo de promoción en su trabajo, mayor facilidad a la hora de obtener financiación para su investigación, etc. Lógicamente, este reconocimiento aumentará cuantas más personas tengan acceso a la publicación. Así mientras para el autor literario las copias o reproducciones de forma ilegal de sus obras suponen un cierto perjuicio a sus intereses, para el autor científico dichas reproducciones suponen una mayor visibilidad de su trabajo, que se traducirá en un mayor reconocimiento internacional.” Se teme que esta situación nos suma en unos años en el subdesarrollo científico². El círculo vicioso se resume en la venta barata de la materia prima –derechos de los escritos científicos- y la compra del producto manufacturado, a precios desorbitados –especialmente revistas científicas-. Así, se produce un perjuicio económico para las bibliotecas y sus organizaciones, además de uno moral, en sentido general, para la ciencia y su desarrollo. Ante ello, véase la iniciativa de

¹ García Testal, Cristina y Barrueco Cruz, José Manuel. Apuntes sobre edición científica electrónica y derechos de autor. *El profesional de la información*, 1998, vol. 7, nº 4. p. 7-11.

² Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H., y Hilf, E. The green and the gold roads to Open Access. *Nature Web Focus*. 2004. Disponible en WWW: <<http://www.nature.com/nature/focus/accessdebate/21.html>>

la Public Library of Science³ por parte de la comunidad científico-médica, designada como la “revuelta de los científicos”⁴. Cuenta ya con casi 34000 investigadores de 180 países que exigen de los editores científicos que, pasado un periodo de tiempo tras la primera edición, distribuyan libremente sus contenidos, en la línea de la normativa norteamericana que basa los beneficios de los editores en la primera venta⁵. De esta situación, las universidades evolucionan hacia la edición de sus propios materiales para rebajar sus presupuestos y proteger a sus científicos de la cesión indiscriminada de sus derechos. Juan Antonio Millán⁶ se pregunta “¿Y de qué vivirán los editores, si se extiende esta posición? De la primera publicación de una obra”⁷. Los miembros de este movimiento creen que los beneficios actuales de los editores de literatura científica son excesivos. Para Michael Eisen, uno de los pioneros en esta iniciativa, los editores deberían ser “como comadronas; se les paga por su trabajo, y al final del día devuelven el niño a sus padres”. Esta tensión entre industria y ciencia, también apuntada en el White Paper de Cox⁸, quizá produzca la necesaria modificación del actual modelo de comunicación científica.

Sociedad de la información

En esta situación irrumpe la sociedad de la información. La penetración de las tecnologías de la información y comunicación ha provocado un cambio radical en el escenario social, y por ende, en la comunidad científica. Por una parte, ha aumentado la cantidad de información disponible en Internet, aunque en su mayor parte no se refiere al conocimiento científico. La facilidad para hacer pública y accesible libremente cualquier información en Internet ha fomentado que los autores⁹ difundan sus trabajos a través de este medio, convirtiéndose en “editores”, al realizar una primera difusión de su obra, denominado *preprint*, cuando todavía no se ha publicado en medios tradicionales. Sin embargo, en realidad nos encontramos con una doble difusión: individual en la red y en publicaciones colectivas impresas. La difusión individual de trabajos (inéditos o publicados) en la red no resulta un medio muy adecuado para que se

³ *Public library of science*. [S.l.: s.n, 2000]. Disponible en WWW: <<http://www.publiibraryofscience.org/>>

⁴ Fernández Hermana, Luis Ángel. La tarascada de la ciencia. *Enredando*, 11/09/2001, nº 284. Disponible en WWW: <<http://enredando.com/cas/cgi-bin/editorial/plantilla.pl?ident=493>>

⁵ Arms, William Y. *Digital Libraries*. Cambridge (Massachusetts): Massachusetts Institute of Technology, 2000. ISBN 0-262-01880-8

⁶ Millán, José Antonio. Vender vino sin botellas (El nuevo circuito editorial). [Versión ampliada de la ponencia presentada en el Congreso "Las tecnologías de la información y sus desafíos", organizado por España Nuevo Milenio, Valencia, 6 de noviembre del 2000]. [S.l.: s.n.], 21 de diciembre del 2000. Disponible en WWW: <<http://jamillan.com/botellas.htm>>

⁷ Tal y como propugna la tradición Estados Unidos: *fair use, first sale*.

Refrendada también en su legislación, *Digital Millennium Copyright Act, Public Law 105-304*, signed by the President October 28, 1998. Disponible en WWW:

<http://www.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=105_cong_public_laws&docid=f:publ304.105>

⁸ Cox, John. *Publisher/Library Relationships in the Digital Environment. An STM White Paper*. Northants: STM Library Relations Committee, April 1999. Disponible en WWW: <<http://209.41.0.61/vops/a06/999/upload/pics/lrcdoc1504.doc>>

⁹ Swan, A. y Brown, S.N. *JISC/OSI Journal Authors Survey Report*. London: The Joint Information System Committee, 2004 Disponible en WWW: <http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf>

encuentren -incluso con Google- y utilicen; además de no suponer por sí misma una publicación a efectos curriculares. De hecho, los mecanismos para validar la fiabilidad y veracidad de este tipo de datos, basados en la evaluación por pares, no cuentan todavía con un modelo alternativo en la red. Por ello, se mantiene el envío a revistas o congresos como forma de certificar su calidad. Independientemente de que se trate de revistas en formato únicamente impreso o con dos versiones (online e impresa), la función de las revistas sigue siendo avalar la calidad de los trabajos que publica mediante, normalmente, la evaluación por un comité editorial científico. Crawford y Gorman incluyen las condiciones ideales de la oferta de información electrónica, que son:

- “Autenticidad: cada usuario debería tener la seguridad de que el documento que encuentra en la red es el documento auténtico y original.
- Protección de la propiedad intelectual: la protección de los derechos de autor debería estar garantizada en todo documento recuperado.
- Asequibilidad: los costes de acceso y recuperación de los diversos documentos tendrían que ser razonables y no superar los costes de sus equivalentes tradicionales.”¹⁰

Así, las bibliotecas y sus servicios digitales, están actuando, en cierto modo, como intermediarios, al tiempo que como “certificadores” de la calidad de la información que recogen en su colección –calidad referida a la autenticidad y veracidad, no sobre el contenido-.

En definitiva, la forma de difusión es sustancialmente la misma, aunque el canal de comunicación sea de soporte y transmisión diferente.

Con ello pasamos al segundo aspecto, el almacenamiento, conservación y segunda difusión de esos nuevos documentos digitales. Las bibliotecas universitarias han asumido hasta ahora esta función: proporcionar sin ánimo de lucro a su comunidad e incluso a pequeños centros sin tanto presupuesto, el acceso y consulta a los fondos bibliográficos.

En este caso sí pensamos que se ha producido una transformación radical. La aparición de Internet ha modificado los hábitos de los consumidores de la información. Usuarios cada vez más exigentes que no quieren, bajo ningún concepto, dejar sus despachos para acceder al documento. Con ello, y dejando aparte la costumbre -tan agradable- de comprar en librerías, los usuarios progresivamente han modificado sus hábitos de consumo de información al poder contar con los fondos de las bibliotecas en sus ordenadores personales. Hasta ahora se localizaban las publicaciones en la biblioteca, encargada de conservarlas de manera organizada. Ahora han aparecido nuevas formas de archivar los trabajos en red –al margen de la primera difusión que un autor puede hacer en sus páginas-. Esta distribución sistemática de textos completos publicados o en vías de publicar en revistas es la que está planteando problemas a los agentes productores/intermediarios de la información científica. Por tanto, hay que puntualizar cómo se armonizan las demandas de las bibliotecas y los usuarios con los derechos de la industria editorial. Para ello

¹⁰ Crawford, Walt y Gorman, Michael. *Future Libraries: dreams, madness & reality*. Chicago: American Library Association, 1995. Pérez, Dora, *La biblioteca digital*. Barcelona: UOC, [s.a.]. Disponible en WWW: <http://uoc.terra.es/articles/La_biblioteca_digital.htm>

estudiaremos el papel de las bibliotecas en el ámbito de la comunicación científica en la sociedad de la información¹¹.

No obstante hay que insistir que, aun no teniendo carácter lucrativo, alrededor de la investigación y ciencia se mueven, en algunos casos como ingenierías o medicina, grandes presupuestos, internos y externos, como ya comentamos. Así, el derecho a la información, sin menoscabo del derecho de la industria editorial a sobrevivir, obliga a las bibliotecas con mayor presupuesto a la organización de recursos de información para alimentar la demanda de sus usuarios. Pero la situación de gratuidad ¿puede asegurarse y mantenerse en bibliotecas digitales dados los costos técnicos y los de compra de información y el derecho a su utilización? Ningún editor o proveedor de información puede impedir, mediante la firma de contratos de licencia de uso, las funciones básicas de las bibliotecas: difusión a través de la posibilidad de búsqueda, uso e impresión, reproducción para la investigación y preservación, derecho de cita...; o secuestrar los derechos morales del autor sobre su obra. Por ello, las bibliotecas y los autores deben estar alerta sobre la normativa de propiedad intelectual nacional e internacional¹² a la hora de firmar contratos o licencias. En cuanto a su distribución, para García Testal y Barrueco Cruz, “los usuarios, las bibliotecas y las instituciones educacionales tienen derecho a esperar:

- Que los términos de las licencias no restringirán el uso justo u otros legítimos de las bibliotecas o los usos con fines educacionales.
- Que los trabajos gubernamentales y otros materiales de dominio público estarán disponibles sin restricciones y a un precio que no exceda el coste mínimo de diseminación; y
- Derechos de uso para la educación sin ánimo de lucro tanto si se aplica en el aula como si es un uso remoto cuando la educación deba impartirse a distancia.”

Bibliotecas digitales universitarias

Para estudiar estas cuestiones vamos a caracterizar como bibliotecas digitales diversos tipos de distribución que afectan de forma diferente a los derechos de propiedad intelectual. Los siguientes casos son denominados bibliotecas digitales en universidades, teniendo en cuenta las siguientes limitaciones: la información proporcionada es gratuita en principio para la comunidad universitaria, está distribuida en red, y afecta al documento llamado primario, es decir, original a texto completo:

1. La distribución de información electrónica organizada que se adquiere a los agentes tradicionales en el suministro de contenidos –proveedores de información y editoriales-. Esto se produce, especialmente para la información que circula en publicaciones periódicas.

¹¹ Guedon, Jean-Claude. In Oldenburg's long shadow: librarians, research scientists, publishers and the control of scientific literature. En *Creating the Digital Future: Association of Research Libraries Proceedings of the 138th Annual Meeting, 2001*. Washington: ARL, [2005]. Disponible en WWW: <<http://www.arl.org/arl/proceedings/138/guedon.html>>

¹² Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual. BOE 22-4-1996. Disponible en WWW: <http://www.mcu.es/Propiedad_Intelectual/anexos/LeyProp_Intelectual_mod171.PDF>, y otros convenios internacionales y directivas europeas. Véase *Creative Commons Worldwide: International Commons: Spain*. [S.I.]: CC, [2004]. Disponible en WWW: <<http://creativecommons.org/worldwide/es/>>

2. La organización de contenidos producidos por la propia institución y su comunidad: tesis, cursos multimedia, investigación, congresos... en forma de auto-archivo o archivos institucionales.
3. La reconversión de materiales a través de la digitalización, especialmente los que ya no cuentan con gravámenes derivados de la propiedad intelectual y poseen sus derechos de reproducción. Suelen constituir parte del patrimonio cultural de un país.
4. La selección y organización de recursos de información, bibliográficos o no, que se encuentran en la red a texto completo.

En cuanto al primer y segundo caso (distribución de información adquirida y edición y organización de contenidos propios de las instituciones de investigación), la implementación de bibliotecas digitales tiene dos frentes básicos que afectan a la edición y los derechos de autor:

- 1) En primer término, parte de la colección está formada por la bibliografía especializada, que circula en publicaciones periódicas y es utilizada mayoritariamente por el investigador. En este caso se observa una clara evolución hacia su distribución electrónica por parte de dos agentes: las editoriales de ámbito científico, y los agregadores de información secundaria a través de acuerdos con los primeros introduciéndola en sus bases de datos a texto completo o mediante enlaces a sus sitios web. Las bibliotecas siempre han promovido el uso responsable de la información, acorde al derecho. No obstante, para Gallart¹³, por ahora las bibliotecas prefieren delegar en editores y proveedores de información la adquisición de la información electrónica –caso del Consorci de biblioteques universitaries de Catalunya¹⁴–, con la consiguiente pérdida de experiencia, sujeción a costes impuestos, y en algunos casos injustos¹⁵, de la firma indiscriminada de licencias de uso...
- 2) En segundo término, gran parte de la colección está formada y seguirá estando formada por la bibliografía básica, que se distribuye en forma de manuales monográficos y es utilizada para la docencia. En este sentido, Gallart destaca el papel de la biblioteca en sus respectivas instituciones para ayudar a la creación de bibliotecas digitales corporativas. Dada la tendencia a acudir a la información electrónica, si los manuales se ofertasen de forma remota, los estudiantes disminuirían su presencia en la biblioteca. Desde luego, si las bibliotecas tienen en mente ofrecer los mismos servicios de una biblioteca clásica de forma remota, debe considerarse como imprescindible la digitalización de los textos docentes. Esto atenta a la supervivencia de la industria editorial directamente, de manera que si esta

¹³ Gallart, Núria, The future of copyright management: library perspectives. En: *66th IFLA council and general conference. Jerusalem, Israel, 13-18 august 2000*. Canadá: National Library, 2000. Disponible en WWW: <<http://www.ifla.org/IV/ifla66/papers/172-184e.htm>>

¹⁴ Consorci de Biblioteques Universitàries de Catalunya. *Què és la biblioteca digital*. Barcelona: CBUC, abril de 2001. Disponible en WWW: <<http://www.cbuc.es/5digital/5mcdigital.htm/>>

¹⁵ Consorci de Biblioteques Universitàries de Catalunya. *Porqué algunas bibliotecas y consorcios están pagando demasiado por la información electrónica: Declaración del Consorcio de Bibliotecas Universitarias de Cataluña sobre los modelos de precios aplicables a bibliotecas y a consorcios de áreas no anglosajonas*. Barcelona: CBUC, abril de 2001. Disponible en WWW: <<http://www.cbuc.es/5digital/55docs.htm>>

tendencia continúa, habrá de plantearse el ofertar estos textos sistemáticamente en formato electrónico. En este caso, desde el terreno de la edición tradicional en España se han dado escasos pasos hacia su digitalización. Por ello, bibliotecas o productores de información deberán afrontar su digitalización, aún siendo costoso y fuera de sus funciones, tal y como han sido entendidas hasta ahora. No obstante, la evolución hacia la formación remota en las universidades presenciales está produciendo, de forma decidida, manuales multimedia para cursos concretos, que generalmente se realizan con cargo a la propia universidad. Así, aunque para la industria la edición académica no suponga la mayoría de su nicho de negocio, ha de tener en cuenta que perderá la carrera en pro de los propios productores de la información. En este sentido, pueden consultarse los informes del grupo de trabajo de la California Digital Library sobre los E-book¹⁶, donde describe algunos de los programas que permiten los micropagos por la información, como WebBuy de Adobe.

En cuanto al segundo y tercer caso (organización de contenidos propios y digitalización de materiales patrimoniales), las bibliotecas tradicionales asumen un papel pionero, pues participan en la implementación de bibliotecas digitales en sus instituciones.

a) En este sentido cabe destacar las iniciativas de implantación de los repositorios de información producida por los propios universitarios. Podemos encontrar archivos de tipo temático y archivos de tipo institucional, como veremos con mayor detalle. En ambos casos, al ser el autor quien autoarchiva los trabajos, no suele atentar contra los derechos de los editores de las revistas.

b) Por otra parte, con los escasos medios que se dedican a las bibliotecas en España, se han comenzado proyectos de digitalización que garantizan el respeto a los derechos de autor y su distribución gratuita –por ejemplo, el proyecto Dioscórides¹⁷-. Tendencias que, quizá con el tiempo, veremos convertirse en nichos de negocio para la industria editorial, al igual que el desarrollo de Internet se ha convertido en una opción de mercado más allá de la comunicación científico-técnica. Para Arms, los productos comerciales no siempre se adecuan a las necesidades de las bibliotecas, obligándolas a experimentar en el campo de las nuevas tecnologías¹⁸.

El cuarto caso, organización de recursos de calidad que circulan gratuitamente por la red, lo podemos encontrar ya en la mayoría de webs de las bibliotecas universitarias. Dado que se trata de información distribuida gratuitamente y de reconocida calidad –pues en caso contrario las bibliotecas no la organizarían-, cada vez más se reconoce su valor y facilidad de acceso frente a los materiales impresos. Estas recopilaciones dan lugar a una clase de bibliotecas digitales, en este caso como repositorios de información de calidad. Este tipo de

¹⁶ Coyle, Karen, et al. *REPORT Ebook Task Force*. California: California Digital Library, Joint Steering Committee for Shared Collections, 2001. Disponible en WWW: <<http://www.cdlib.org/libstaff/sharedcoll/jsc/ebook/>>

¹⁷ Biblioteca digital Dioscórides. Biblioteca Histórica Marqués de Valdecilla. Madrid: UCM, 2000. Disponible en WWW: <<http://www.ucm.es/BUCM/foa/dioscorides.htm>>

¹⁸ Arms, William Y. *Digital Libraries...*

biblioteca digital no afecta a los derechos de propiedad intelectual directamente, ya que, con independencia de que una biblioteca los organice o no, estos materiales ya existen a texto completo en la red. Hay que insistir en que progresivamente aumenta la información de calidad disponible en Internet, pues la mencionada revuelta de los científicos ha comenzado a modificar las vías tradicionales de comunicación científica y la publicación electrónica comienza a reconocerse en los *curricula* de investigación.

Para cualquiera de estos cuatro casos, pero especialmente del primero al tercero, Gallart insiste en la necesidad de conocer cuáles son los derechos y condiciones de difusión que las bibliotecas poseen sobre sus colecciones y cómo afecta la legislación a sus actividades. Pero conocer los derechos y condiciones de difusión de la información no es tarea fácil. Especialmente lo resalta Parry cuando se trata de procesos de digitalización y difusión de obras no recientes¹⁹. Esto viene señalado claramente por Gallart o Harris²⁰, este último abogado especializado en propiedad intelectual, quien ofrece quince pasos básicos para proteger los derechos de las bibliotecas. No obstante, la Digital Millennium Copyright Act de 1998 exime a las bibliotecas digitales de toda responsabilidad por malos usos de la información mientras cumplan ciertos requisitos como la advertencia legal, la obediencia a las autoridades, la información de los malos usos que conozcan... Teniendo en cuenta que Estados Unidos siempre ha sido pionero en la aplicación de las nuevas tecnologías en las bibliotecas, es sorprendente que en Europa se pongan tantas trabas legales a la implementación de estas bibliotecas digitales.

Open Archives Initiative

En este sentido, uno de los hitos más relevantes de los últimos años es el desarrollo del movimiento Open Archives Initiative²¹. Derivado de un intento de promover el libre acceso al conocimiento científico, actualmente ha evolucionado a una red de archivos que proporcionan los datos, y servicios que los consultan, independientemente del software utilizado para almacenarlos o del soporte en que se registren. Ya García Testal y Barrueco preconizaban ciertos cambios en bibliotecas digitales: un pago mayor por ciertos contenidos, de los que se controlará exhaustivamente su uso, y unos servicios sobre ellos más baratos. Así, el valor mayoritario no resulta del uso de los contenidos - pues sólo unos pocos serán de alto coste-, sino de la utilización de los servicios que se ofrezcan sobre ellos –selección, garantía de autenticidad, es decir, de la información fidedigna acerca de las fuentes y las corrientes futuras-.

En este contexto, el archivo de los trabajos científicos en bibliotecas digitales de libre acceso es el primer paso para disponer del conocimiento. Como hemos

¹⁹ Copyright and intellectual property rights. En: *Virtually new: creating the digital collection. A review of digitisation projects in local authority libraries & archives*. Prepared by Information North Consultants to the Review ; Report by David Parry. London: Library and information commission, 1998. Disponible en WWW:

<<http://www.ukoln.ac.uk/services/lic/digitisation/intro.html>>

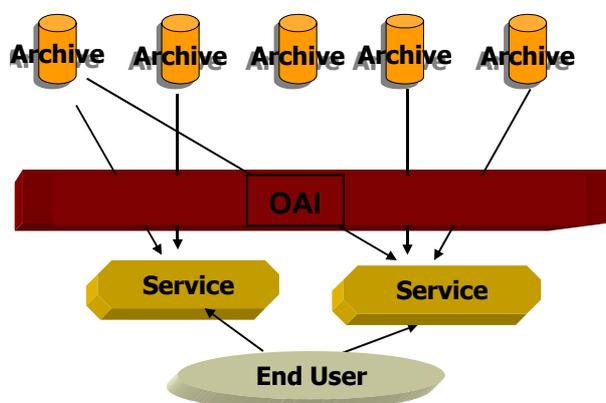
²⁰ Harris, Lesley Ellen. Finding Your Way Out of the Copyright Maze. *Computer in libraries*, June 1998, vol. 18, nº 6. Disponible en WWW:

<<http://www.infotoday.com/cilmag/jun98/story1.htm>>

²¹ *Open Archives Initiative*. [S.l.: OA, 2000-]. Disponible en WWW:

<<http://www.openarchives.org/>>

adelantado, en OAI son los autores quienes se encargan de cargar los trabajos en cada una de las bibliotecas digitales, lo que se denomina autoarchivo²². Estas bibliotecas, llamadas archivos o *data providers* en el contexto de OAI, son definidas por el grupo que implementa el software, por lo que cada contenido introducido o aplicación utilizada para la gestión puede ser diferente. Pero dentro de la diversidad este movimiento tiene una base común, pues en OAI los datos, es decir las contribuciones a publicaciones científicas, siempre se almacenan en archivos, y con independencia de los datos se establecen servicios. Éstos son los que proporcionan diversas acciones que añaden el valor a los simples datos (por ejemplo consultando simultáneamente varios archivos –OAIster²³-, o enlazando la bibliografía referenciada por los trabajos –OpCit²⁴-. El siguiente gráfico muestra cómo se articula, indicando que los servicios no tienen por qué “atacar” –consultar, estar referido, etc.- los datos de todos los archivos:



Fuente: Modificación de ReDIF Barrueco, JM. Experiencias cooperativas en archivos abiertos en Ciencias Sociales. *IV Semana de la ciencia*, Madrid 2004 (modificado)

El derecho como disciplina

Como ejemplo de la situación que queremos ilustrar hemos realizado un análisis referido a la disciplina del derecho. Hay que tener en cuenta que este movimiento es de carácter internacional, mientras que por las características del sector que se contempla muchos de los trabajos serán publicados a escala nacional. Pero independientemente de esta circunstancia, también el acceso abierto a la bibliografía sobre derecho forma parte de la situación que hemos mostrado. El análisis de varios aspectos nos ayudarán a sistematizar este escenario:

- a. En primer lugar las publicaciones que editan y distribuyen los trabajos bajo suscripción, así como los editores que las producen.

²² Harnad, Stevan. The self-archiving initiative. *Nature WebDebates* 2001, nº 410. p. 1024-1025. Disponible en WWW: <<http://www.nature.com/nature/debates/e-access/Articles/harnad.html>>

²³ *OAIster... find the pearls*. Michigan: University, 2003. Disponible en WWW: <<http://oaister.umdl.umich.edu/oaister/>>

²⁴ *OpCit: the open citation project. Reference Linking and Citation Analysis for Open Archives*. Southampton: University, 2004. Disponible en WWW: <<http://opcit.eprints.org/>>

- b. En segundo lugar, las materias primas, que siguen siendo los artículos científicos que un autor crea, pero también otro tipo de producto que no ve la luz en revistas como son informes técnicos, presentaciones...
- c. En tercer lugar, los archivos que lo almacenan gracias al esfuerzo del autor. De ellos hay que conocer el software, el contenido y el vocabulario para archivar la información.
- d. Por último, los servicios que se ofrecen sobre los datos, para cuya ilustración se mostrará METALIS.

a. Publicaciones

Poco nuevo hay que decir sobre las publicaciones, pues siguen asumiendo sus funciones tradicionales. El único cambio reseñable es que pueden tener una versión online distribuida, generalmente, desde los sitios web de la propia editorial o adquirida y servida por las bibliotecas.

Pero la modificación del entorno ha producido una flexibilización en la gestión de los derechos exclusivos de difusión y reproducción de los trabajos. Algunos editores permiten a los autores autoarchivar –por tanto, difundir- sus trabajos. Esta situación, entre otras cosas, es lo que pretende conocer para todas las disciplinas y a escala internacional el proyecto Sherpa²⁵, heredero de RoMEO. Esta iniciativa cooperativa, patrocinada por CURL -Consortium of research libraries in the British Isles- y JISC -Joint Information Systems Comitee-, ofrece una base de datos de títulos de revista. Utilizan unos códigos de colores para indicar si la revista no permite el almacenamiento de los trabajos que publica en archivos de acceso abierto (gris), si permite el *preprint* todavía no evaluado (verde pálido) o si permite el *postprint* ya evaluado por *referees* (verde). Para todos los títulos de revista introducidos, Harnad²⁶ realizó un estudio sobre 10673 revistas y 88 editores:

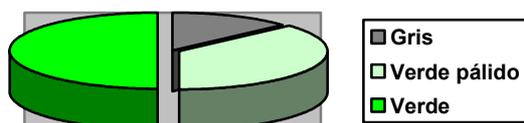
Green light to self-archive:	Journals	%	Publishers	%
	10,673	(100%)	88	(100%)
Neither yet	1,793	17%	37	42%
Preprint	3,253	+30% (=83%)	7	+8% (=58%)
Postprint	1,772	+17% (=53%)	14	+16% (=50%)
Postprint and Preprint	3,855	36%	30	34%

Fuente: RoMEO Directory of Publishers who have given their green Light to self-archiving.
<http://www.dlib.org/dlib/june04/harnad/06harnad.html>

²⁵ SHERPA. Nottingham: University of Nottingham, 2004. Disponible en WWW: <<http://www.sherpa.ac.uk/>>

²⁶ Harnad, S. & Brody, T. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals, *D-Lib Magazine* June 2004, vol. 10, nº 6. Disponible en WWW: <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>

El resultado era concluyentemente favorable al autoarchivo, lo que vamos a corroborar con las revistas referidas al Derecho. A modo de ejemplo se ha interrogado por el término LAW en la base de datos RoMEO, dando 112 publicaciones en total. De ellas 15 no permiten el autoarchivo (grises), 41 lo permiten con restricciones y 56 lo permiten totalmente.



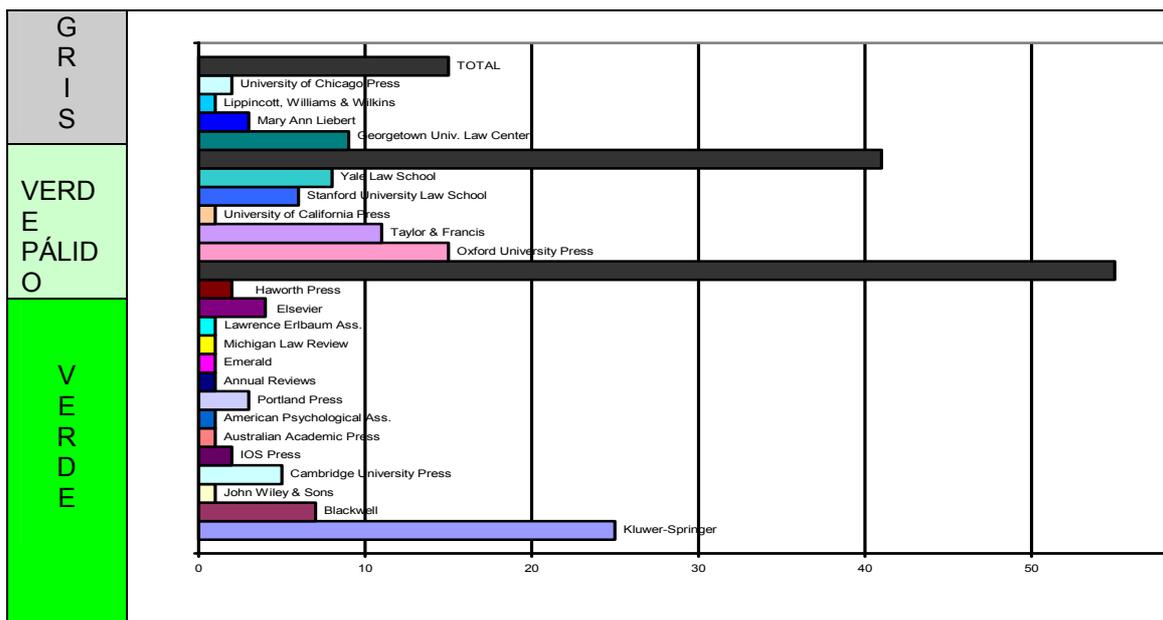
Fuente: elaboración propia con datos de RoMEO

No obstante, existen excepciones en cada categoría. Éstas, para el verde pálido conforman dos grupos principalmente. Los que permiten el *preprint*, pero no el *postprint*, que son la mayoría (80,48%) -dentro de este grupo, existen algunas revistas que permiten el *preprint* antes de la publicación solamente (18,18%)-. El otro grupo (19,52%) permite el *postprint* bajo la condición de que se haga un año después de la publicación del archivo. La mayoría, pues permiten el *preprint* o siguen la costumbre estadounidense de liberar el derecho de difusión al cabo de un año.

La gran mayoría de los títulos clasificados en verde obedecen a la condición de *preprint* y *postprint* libres (78,57%). Sin embargo, existen una serie de excepciones en este grupo de libre acceso: algunos no permitiendo el *preprint* y condicionando el *postprint*. Por ejemplo, exigiendo el consentimiento del editor para utilizar publicaciones en PDF (tres casos), la obtención de permiso antes de la reutilización, o la limitación al uso de la versión publicada (dos casos).

En definitiva, sólo el 50 % de los títulos de revista en esta disciplina permite al autor autoarchivar sus trabajos, cifra idéntica a la que Harnad ofrece para todos los títulos de revista.

Los datos extraídos también permiten clasificar la postura ante el autoarchivo de cada uno de los editores de los títulos de revista que recoge RoMEO. Los quince títulos que no lo permiten están publicados por cuatro organizaciones, los cuarenta y uno que permiten alguna forma de autoarchivo están publicados por otras cinco, y el grueso, 56, los publican catorce editores. Resulta por tanto un total de 23 editores que pueden dividirse entre organizaciones de tipo universitario -8-(privadas en su totalidad) y editoriales comerciales -15-. Desde esta nueva perspectiva hay que destacar que, en términos generales, las organizaciones que no permiten autoarchivar los trabajos tienen carácter universitario -privado-, mientras que las grandes editoriales comerciales internacionales como Kluwer-Springer, Blackwell, Emerald, Elsevier o Haworth sí lo permiten, como muestra el siguiente gráfico:



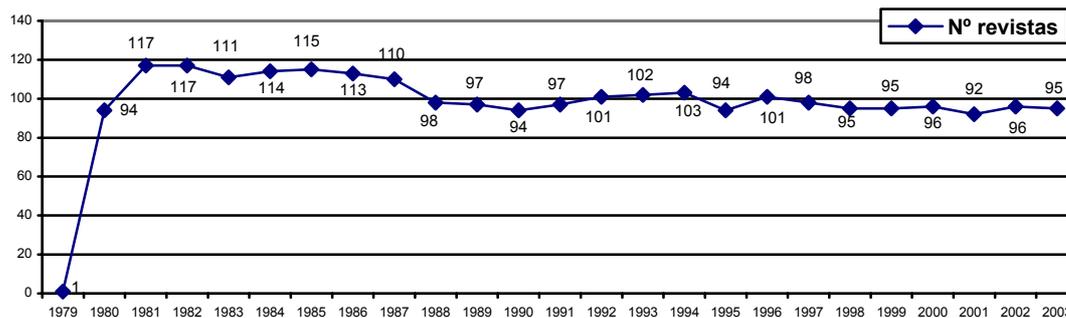
Su significado es más que elocuente. Quizá se pueda considerar que la presión para que las editoriales privadas permitan el autoarchivo está surtiendo efecto, al menos en la disciplina analizada.

La evolución cronológica la tomamos de un estudio que realizó Hajjem²⁷, de la Universidad de Québec, para el análisis de impacto de los artículos de libre acceso en web. Para ello programó un robot que rastrea automáticamente los artículos científicos en web y calcula el número de citas que reciben para extraer la siguiente información. Los objetivos concretos son:

calculer le pourcentage d'articles publiés sur le Web, le rapport entre le nombre de citations des articles publiés sur le Web et les articles non publiés, calculer l'avantage de l'impact de citations, dessiner les graphiques identifiant la variation des divers variables (nombre total d'articles, nombre d'articles publiés sur le Web, avantage de citations) avec le temps.

Si bien no está claramente especificada la metodología de recogida de datos, los resultados, aun siendo estimativos, son útiles para nuestros fines. Se observa un máximo de 117 en 1981 y 1982, que en 2003 cae a 95 títulos de revista de derecho que contienen artículos de libre acceso sobre derecho.

²⁷ Hajjem, Chawki. *Dossier du projet. Robot spécialisé dans la recherche des articles scientifiques sur le web*. Québec : Université, 2004. Disponible en WWW: <<http://www.crsc.uqam.ca/lab/chawki/projet.pdf> 3-2-2005>

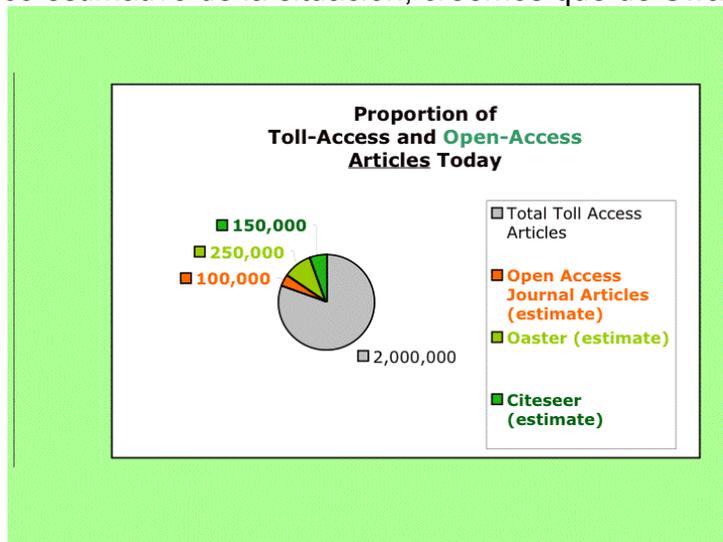


Fuente: Elaboración propia a partir de datos de Hajjem, Chawki. Laboratoire de recherche en Sciences Cognitives <http://www.crsc.uqam.ca/lab/chawki/ch.htm> 3-2-2005

Para interpretar correctamente el alcance de estos datos tenemos que reseñar que la introducción de la información bibliográfica en bases de datos suele ser relativamente lenta –excepto en los servicios de alerta-. Por ello, se produce siempre una mayor circulación de artículos, títulos o impacto de las citas de los documentos con una cierta obsolescencia.

b. Artículos de revista

Aún más difícil es cuantificar esta segunda entidad, el número de artículos que se encuentran en libre acceso. En términos generales, podemos partir del siguiente gráfico estimativo de la situación, creemos que de Swan y Brown²⁸:



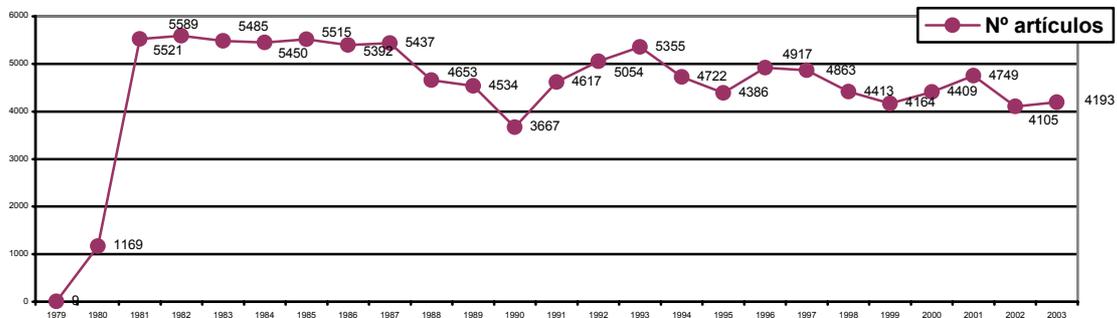
Fuente: *Proportion of Toll-Access and Open-Access articles today.* <http://www.ecs.soton.ac.uk/~harnad/Temp/Romeo/romeosum.html>

Puede apreciarse que la cantidad de artículos disponibles en libre acceso es escasa frente al resto. No obstante, en estos momentos se trabaja con datos estimativos, más aún cuando no está clara la metodología con la que realizar un análisis estadístico que responda a la realidad.

²⁸ Swan, Alma y Brown, Sheridan. Authors and open access publishing. *Learned Publishing*, 1 July 2004, vol. 17, nº 3, p. 219-224

Tomando prestados de nuevo los datos de Hajjem²⁹ podemos comparar la situación del derecho. Por un lado presenta una evolución cronológica anual desde el año 79 hasta el 2003. Por otro permite la comparación con otras disciplinas.

En su evolución cronológica se aprecia una franja de máximos que se mantiene desde 1981 a 1987.



Fuente: Elaboración propia a partir de datos de Hajjem, Chawki. Laboratoire de recherche en Sciences Cognitives <http://www.crsc.uqam.ca/lab/chawki/conclusion.xls> 3-2-2005

En segundo término sus datos permiten la comparación con las otras disciplinas. Para cada área de conocimiento ofrece la media de artículos ofrecido en libre acceso por revista y distribuye los resultados en tres frecuencias, aproximadamente la primera comienza de 0 a 99, la segunda de 100 a 203 y la última de 204 en adelante. Para el primer estrato la media es de 70,96 artículos de libre disposición por revista, para el segundo 133,13 y para el tercero 262,05. Y aquí viene lo interesante: la disciplina del derecho aparece con el valor mínimo del estudio, 45,94 artículos de libre acceso por revista; mientras que la máxima la tiene medicina clínica, con 231,78. Es obvio que la diferencia de casi doscientos ítem sitúa en planos absolutamente diferentes a una y otra área de conocimiento. Medicina es una disciplina con gran tradición internacional, al tiempo que cuenta desde 1999³⁰ con PubMed Central, biblioteca digital de literatura científica de carácter centralizado.



Fuente: PubMed Central. <http://www.pubmedcentral.nih.gov/>

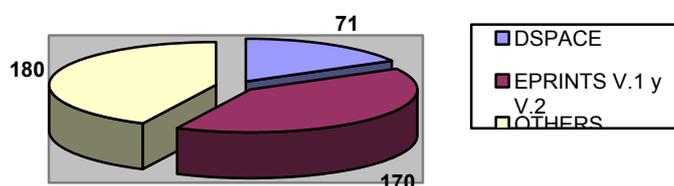
²⁹ Hajjem, Chawki. *Dossier du projet...*

³⁰ Harnad, Stevan. Free at Last: The Future of Peer-Reviewed Journals. *D-Lib Magazine* December, 1999, vol. 5, nº 12 y E-Knowledge: Freeing the Refereed Journal Corpus Online. *Computer Law & Security Report*, 2000, vol. 16, nº 2. p. 78-87. Disponible en WWW: <<http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.scinejm.htm>>

c. Archivos: software, contenidos y vocabularios

El final del apartado anterior nos conduce al penúltimo aspecto a estudiar: qué archivos / bibliotecas digitales existen para el área del derecho definidas como de libre acceso y con un protocolo estandarizado. Repositorios de documentos existen de todos los tipos, pero ya hemos comentado que Open Archives Initiative se estructura en dos vertientes: unos archivos que proporcionan los datos y unos servicios que actúan sobre esos datos. De esta manera, la información no se encuentra totalmente centralizada, siguiendo la filosofía del protocolo de Internet TCP/IP. Para poder compartir los datos proporcionados por los archivos éstos se han de haber implementado con el protocolo OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting), de manera que los servicios recolecten los metadatos descriptivos de cada uno de los documentos.

Sin querer entrar en detalles, en la actualidad existen varios softwares que utilizan este protocolo, cuyo registro se encuentra en las páginas de uno de ellos³¹. Ofrece un listado de todos los archivos registrados, 421, con independencia del programa utilizado. Entre ellos aparece EPrints³², desarrollado por la Universidad de Southampton, que cuenta con el 40 %; también DSpace³³, desarrollado por una alianza entre el MIT-Massachusetts Institute of Technology y HP, con 17 %; y en tercer lugar el resto, como CDSWare, FEDORA, ARNO, u otros sin identificar por la fuente utilizada.



Fuente: Elaboración propia con datos de <http://archives.eprints.org/?page=all>

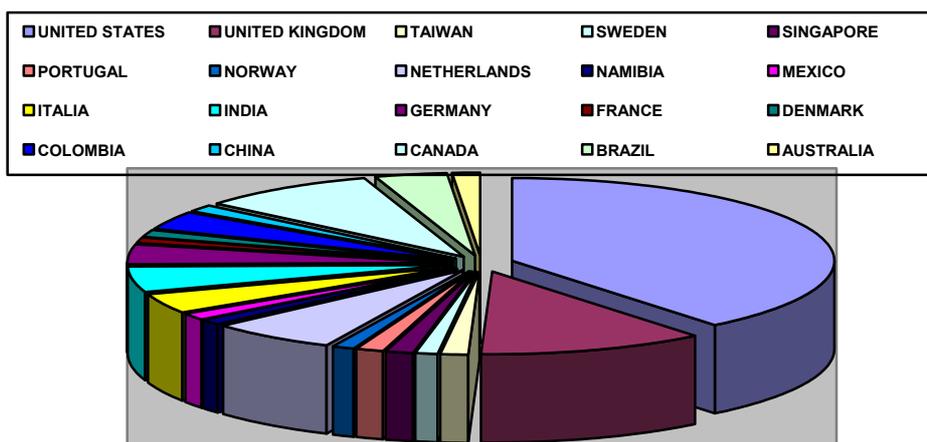
Por su parte, el website de DSpace ofrece un listado de instituciones que lo utilizan³⁴, que constituyen 79 y se dividen de la siguiente forma por países, siendo mayoritarios los Estados unidos de América y Reino Unido.

³¹ *Institutional Archives Registry – Home*. Southampton: University, 2002-. Disponible en WWW: <<http://archives.eprints.org/?page=all>>

³² *eprints.org sites at Southampton serving Open Archives*. Southampton: University, 2002-. Disponible en WWW: <<http://www.eprints.org/>>

³³ *DSPACE*. Massachussets: MIT Libraries, 2003-. Disponible en WWW: <<http://dspace.org/>>

³⁴ *DspaceInstances*. Massachussets: MIT Libraries, 2003-. Disponible en WWW: <<http://wiki.dspace.org/DspaceInstances>>



Fuente: Elaboración propia con datos de DspaceInstances.
<http://wiki.dspace.org/DspaceInstances>

Podemos comparar estos datos con los que ofrece el registro de Eprints, que incluye todos los archivos, independientemente del programa utilizado. El total analizado suma 421 archivos, que organizados por continentes muestran un predominio aplastante de América del Norte y Europa –véase la tabla-. Tomando otros criterios de análisis, como por ejemplo el idioma, USA, más Gran Bretaña, más Australia, más parte de Canadá supondrían más del 50% de los archivos, concretamente 224.

PAÍS	CONTINENTE	NÚMERO
United States	AMÉRICA	127
Brazil		19
Mexico		3
Colombia		3
Chile		2
Argentina		1
Peru		1
Canada		28
TOTAL AMÉRICA		184
United Kingdom		EUROPA
Germany	37	
France	19	
Italy	16	
Netherlands	16	
Sweden	11	
Portugal	5	
Belgium	5	
Hungary	4	
Denmark	4	
Finland	4	
Switzerland	3	
Norway	3	
Austria	3	
Ireland	2	
Greece	2	
Slovenia	1	
Croatia	1	
Spain	1	
Turkey	1	
Russian Federation	1	
TOTAL EUROPA	193	
India	ASIA	6
Pakistan		1
China		4
Japan		4
Taiwan		2

Singapore		1
Israel		1
TOTAL ASIA		19
South Africa	ÁFRICA	4
Namibia		1
TOTAL ÁFRICA		5
Australia	OCEANÍA	15
TOTAL OCEANÍA		15

Fuente: Elaboración propia con datos de <http://archives.eprints.org/?page=all>

En segundo lugar hemos de estudiar el contenido de estos archivos y los vocabularios que utilizan para implantarlos. En el mundo de los trabajos científicos suele existir una especialización temática, de manera que las bases de datos o bibliotecas digitales abarcan un solo tema. En este sentido, también en Internet se implantaron lo que se conoció como pasarelas temáticas, algunas de las cuales condujeron a bibliotecas digitales de éxito como EconPapers-Economics Working Papers³⁵ o E-LIS-Eprints in Library and Information Science³⁶. Este tipo de archivos podrían denominarse temáticos. Pero además de esta vertiente temática también existen otro tipo de bibliotecas que utilizan el protocolo OAI-PMH, que podríamos llamar repositorios institucionales. La ventaja de aplicar este protocolo a estos archivos que pueden tener una cantidad reducida de datos es que un servicio puede consultar simultáneamente varios de ellos.

Concretamente, en el campo del derecho los únicos archivos identificados en el registro de Eprints han sido tres, y utilizan todos este software: Duke Law Faculty Scholarship Repository³⁷, School of Economics and Commercial Law de la Universidad de Göteborg³⁸ y Pace Law Prints³⁹ (dos norteamericanos y uno sueco). La metodología para clasificarlos como tales ha sido somera, tomando como indicador únicamente el término “law” en título. Esto significa que otros archivos podrían contener trabajos en la disciplina aunque no lo reseñen en título –si por ejemplo la Universidad Politécnica de Valencia implementase un archivo para los trabajos de sus profesores contendría todas las materias que se imparten en esta universidad-. Pero esta simple indicación es útil para definir su contenido.

A modo de ejemplo ofrecemos un resumen de sus características –entre ellas el vocabulario utilizado para su indización- y, como indicamos, observamos que la cuantía no es excesiva teniendo en cuenta que el archivo sueco incluye tesis y tesinas: 2477 trabajos almacenados aproximadamente.

Universidad	Periodo	Vocabulario	Nº trabajos	Año de implantación
Pace	1986	Library of Congress	24	
Duke -Lawarchive	1951	“	953	
Göteborg - Electronic Publishing Center		DIVA-Digital Scientific Archive	1500	2003

Fuente: elaboración propia

³⁵ *EconPapers*. Orebro (Suecia): University, [1998-]. Disponible en WWW: <<http://econpapers.repec.org/>>

³⁶ *E-LIS: Eprints in Library and Information Science*. Milan: CILEA, 2003- . Disponible en WWW: <<http://eprints.rclis.org/>>

³⁷ *Faculty Scholarship Repository*. Durham: Duke Law, [2000-]. Disponible en WWW: <<http://eprints.law.duke.edu/>>

³⁸ *Electronic Publishing Center*. Göteborg (Suecia): University, [2003-]. Disponible en WWW: <<http://www.handels.gu.se/epc/>>

³⁹ *Pace Law Prints*. New Cork: Pace Law School, [2000-]. Disponible en WWW: <<http://eprints.law.pace.edu/>>

d. Servicios

Como punto final, esta constelación de datos necesita servicios que hagan un uso eficiente de la información almacenada con el protocolo OAI-PMH. Como en cualquier biblioteca digital es posible consultar cada uno de los archivos independientemente, pero el valor que añade este protocolo es la posibilidad de ser interrogados simultáneamente por un servicio o por motores de búsqueda como Google, de manera que se trata de bases de datos visibles (al contrario que los catálogos tradicionales de bibliotecas o la mayoría de las bases de datos, que forman parte de la Internet invisible/profunda⁴⁰).

El servicio de consulta más conocido es OAIster⁴¹ -que sirve unos cinco millones de registros de unos quinientos archivos-, pero el más reciente es de noviembre de 2004 para el campo de la documentación -9 archivos con un número mucho menor de trabajos-. Para más detalles, el website de openarchives⁴² ofrece una lista de proveedores de servicios, tanto de consulta, como el presentado, como otras experiencias para medir el impacto, interrelacionar trabajos...

Conclusión

El establecimiento de bibliotecas digitales es imparable hoy día. Tan imparable como imprescindible para controlar la información de calidad que circula por la red. El desarrollo de archivos y servicios basados en la filosofía de Open Archives Initiative puede contribuir a la difusión de la ciencia y la universalización del conocimiento de una manera eficiente.

La utilización de metadatos para archivar los datos que sean recuperables por otros sistemas de interrogación aporta dos claras ventajas. Por un lado, permite una visibilidad mucho mayor⁴³ de la información de los autores. Y por otro, al tratarse de información de calidad repercute en la preservación del acervo científico de cada país, con las implicaciones patrimoniales que conlleva. En definitiva resulta un sistema eficiente para los objetivos de las partes implicadas: autores, bibliotecas y editores.

Concretamente, el análisis realizado presenta un panorama estrecho todavía para el campo del derecho -escasos archivos, escasos artículos...-. Sin embargo, es alentadora la postura de los editores de las revistas científicas en este campo frente al autoarchivo, permitiéndolo la mayoría de ellos. La evolución de los editores ya fue vaticinada, pues mediante la tecnología -marcando textos para rastrear copias ilegales⁴⁴- se consigue que la copia legal sea más eficiente que la ilegal.

⁴⁰ Peset, M. Fernanda, Albiñana, Ricardo, Morales, Silvia. Internet invisible. Un recurso terciario en la red. *Information world en español*, 2000, vol. 9, nº 7-8. p. 19-22

⁴¹ OAIster... find the pearls....

⁴² *Registered Service Providers*. [S.l.: OA, 2000-]. Disponible en WWW: <<http://www.openarchives.org/service/listproviders.html>>

⁴³ Harnad, S. y Brody, T. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, June 2004, vol. 10, nº 6. Disponible en WWW: <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>

⁴⁴ García-Sicilia, Francisca y Alba, Mercè. Distribución electrónica y derechos de autor. En: *7es Jornades Catalanes de Documentació*. Barcelona: SOCADI-COBDC, 1999.

Por último, esperamos que esta contribución ayude a conocer estas iniciativas en nuestro país y anime a otras disciplinas e instituciones a desarrollar los archivos que consideren necesarios. Y, sin más, un abrazo a nuestro padre, cuya capacidad de adaptación tecnológica es de sobra conocida.