

XIII Jornadas de Gestión de la Información

Servicios, Ética, Derechos humanos, Identidad y Compromiso social

E-ciencia y la información como bien público, algunas propuestas.

Lourdes Pérez González. Biblioteca Universitaria, Universidade de Santiago de Compostela.
mlourdes.perez@usc.es

Resumen: Información y conocimiento son las bases para el buen avance de la humanidad, los límites a su acceso y circulación tienen importantes consecuencias en relación al desarrollo de la ciencia, la tecnología y su posición de éstas con la sociedad y los derechos humanos. En este trabajo se darán a conocer algunos de los agentes implicados en contexto de protección (monopolio-mercantilización) / liberalización (acceso abierto-información como bien público) de la información y sus posicionamientos, poniendo como ejemplo de estos últimos proyectos como *CODATA*, *Science Commons*, *Open Data*, centrados crear herramientas para el acceso abierto a los datos científicos, prioridad dentro de las nuevas ciberinfraestructuras de información y la e-ciencia. Partimos del concepto de ciberinfraestructura dado por la *Nacional Science Foundation*, NSF: «una nueva forma de cultura científica que se sustenta en una infraestructura tecnológica de alto nivel, a través de la cual se intenta dar apoyo a nuevos mecanismos de colaboración, basados en el acceso a una ingente cantidad de datos, recursos de información interpretados y reutilizados por potentes herramientas de observación, visualización y simulación. Es un medio que permite el acceso y circulación del conocimiento distribuido, en el que colaboran y se comunican diferentes comunidades y disciplinas rompiendo fronteras culturales, geográficas y temporales. Su fin es crear una base de datos global, cuya sostenibilidad dependerá enteramente del trabajo y esfuerzo coordinado de la interoperabilidad, el consenso normativo en políticas de propiedad intelectual y la consideración y consolidación de la preservación de la ciencia como bien público. Conseguir una e-ciencia accesible para toda la comunidad científica, requiere la acción conjunta de instituciones, autores y profesionales de la información, con el fin de poner en práctica dentro de las estrategias de información políticas de sostenibilidad basadas en la ética del acceso abierto al conocimiento científico».

Palabras clave: *Open data*, e-ciencia, ciberinfraestructuras, *Science Commons*, acceso abierto, *Open Data*, *CODATA*

Abstract: Information and knowledge are the basis for good progress of humanity, the limits on their access and circulation have important implications regarding the development of science, technology and its position with society and human rights. This paper will introduce some of the agents involved in context of protection (monopoly-commoditization) / liberalization (open access as a public-information) of information and their positions, citing the example of recent projects such as *CODATA*, *Science Commons* or, *Open Data*, focused to create tools for open access to scientific data, priority within the new information cyberinfrastructure and e-science. We based on the concept of cyberinfrastructure given by the National Science Foundation, NSF: «a new form of scientific culture is based on a high-technology infrastructure, through which attempts to support new collaborative mechanisms, based on access to an immense amount of data, interpreted and reused information resources by powerful tools of observation, visualization and simulation. It is a medium that enables access and circulation of information distributed, which collaborate and communicate different communities and disciplines breaking cultural, geographical and temporal barriers. Its purpose is to create a global database, whose sustainability depends entirely on the work and effort of interoperability, the normative consensus intellectual property policy and consideration of the preservation and consolidation of

science as public good. Achieving an e-science accessible to the entire scientific community, requires joint action by institutions, authors and information professionals, in order to implement strategies within the sustainability policy information based on the ethics of open access to scientific knowledge».

Keywords: Open data, e-science, cyberinfrastructure, Science Commons, Open Access, Open Data, CODATA.

Introducción

Las nuevas tecnologías de la información y comunicación han transformado la forma en que se crea, distribuye y almacena la información. Alrededor de estas transformaciones han acaecido otras que, aunque menos visibles, han afectado a la circulación de ideas y conocimiento. Nos referimos a los endurecimientos en torno a la protección de la propiedad intelectual en un mundo de economía definitivamente «global». Patentes, marcas, derechos de autor, se han posicionado del lado del mercado, como lo prueba el proteccionismo de la propiedad intelectual mediante diferentes acuerdos multilaterales liderados por los *TRIPS* (1994) que han supuesto la anulación de algunas de las excepciones en relación a la educación, transferencia tecnológica y difusión de la ciencia presentes hasta hace poco en las Convenciones de Berna para los derechos de autor y en la de París para la Propiedad intelectual (Sell, 2007; Varela Orol, 2005).

De esta manera, la academia, la ciencia y todos los agentes implicados en la difusión y acceso a la misma están en una situación de precariedad y desajuste frente al mercado de la información. Ciertamente es que, en la comunicación científica tradicional, estos desequilibrios se podían ir subsanando a través de la cooperación, la digitalización y la creación de repositorios institucionales. Pero el escenario ahora es ya cualitativamente diferente. Nos hallamos, de forma irreversible, ante el nuevo mundo de la e-ciencia. Y es preciso neutralizar, y anular en lo posible, los desequilibrios, desigualdades y límites en el acceso a la información mediante políticas e instrumentos muy precisos o principios de acceso a la información claros y concretos.

Si la e-ciencia promete permitir la colaboración ‘distribuida’ de forma global así como el acceso a la investigación científica a través de la red, al mismo tiempo ésta se encuentra constreñida a las limitaciones derivadas de las normas de propiedad intelectual asociada a los datos, su acceso, uso y preservación para el futuro. Si las normativas de propiedad intelectual suponían en sí una frontera al acceso a los resultados de investigación, la transferencia tecnológica y la comunicación científica en el mundo físico, la situación se agrava dentro del contexto de las ciberinfraestructuras basadas en ambientes colaborativos de diferentes ubicaciones físicas con todas sus barreras institucionales, culturales y legales. Mientras que una de las soluciones ofrecida al problema es la utilización de licencias de «código abierto», lo cierto es que, en ocasiones, la propia composición de las ciberestructuras y sus políticas no permiten adoptar dichas licencias. Es necesario trabajar, por tanto, en políticas de gestión de datos a nivel global que permitan el acceso y la preservación de datos para el futuro.

Los modelos que se están desarrollando e implantando están en la línea de creación de archivos digitales federados y temáticos que alberguen diferentes niveles de colecciónⁱ de datos y en donde se prime la obligatoriedad de depósitoⁱⁱ de aquellos resultados científicos que han recibido financiación pública. Se trata, con estas políticas, de garantizar la preservación de dicha información, al menos, en uno de ellos.

1.1 La importancia de los datos y el acceso a los mismos.

En los últimos años diferentes organizaciones científicas vienen reclamando la necesidad de llevar a cabo el almacenamiento y la preservación de los datos científicos para el futuro, entendiendo que su gestión es una tarea esencial para la consecución de ciberestructuras robustas y sostenibles en información. Su catalogación como bien económico viene justificada por la necesidad de protección de la inversión en investigación, visibilidad de las instituciones, nuevas oportunidades de investigación en el futuro, materia prima y soporte para el conocimiento públicoⁱⁱⁱ.

Esa necesidad de almacenamiento para su preservación en el tiempo para otros usos comienza a ir respaldada en la práctica por nuevas alternativas surgidas en la comunidad científica y académica que reclaman los resultados de las investigaciones como *Open data*. Esto supone trasladar conceptos ‘nacidos’ o surgidos en

movimientos como *GNU, Open Access*^{iv} y modelos de licencias como *Creative Commons*, a los productos de la producción científica. El objetivo es que los resultados de las investigaciones científicas financiadas con fondos públicos, así como el de aquellas financiadas con fines privados pero que no sean explotados por los mismos, puedan ser accesibles a la comunidad académica y científica para su uso y publicaciones sin restricciones de tipo legal determinadas por las actuales normas de propiedad intelectual y derechos de autor.

1.2 Algunas propuestas.

En esta línea, habría que destacar tres de estas propuestas, actualmente las más punteras:

1. *CODATA*
2. *Science Commons*
3. *Pearson*

1.1.1. CODATA

Committe on Data for Science and Technology: Comité interdisciplinar que surge dentro del Comité Internacional para la Ciencia. Entre sus objetivos se encuentra el de estudiar e informar sobre problemas y soluciones en relación al almacenamiento y acceso de los datos científicos.

Desde 1966 ha impulsado la elaboración de estándares y políticas que permiten eliminar cualquier obstáculo social, tecnológico o legal en relación a la colaboración científica y el conocimiento acumulativo. Entre sus distintos informes está *Bits of Power: Issues in Global access to scientific data* (1997)^v, que, pese a estar redactado a finales de los 90, sigue constituyendo la base teórica y programática de informes y políticas similares que se están implantando en el momento presente.^{vi}

1.1.2 Science Commons.

El **Science Commons** se enmarca dentro de los **Creative Commons**, proyecto que surge ante la imperiosa necesidad de dar soluciones al cambio legal en el nuevo modelo de creación, difusión y consumo de ideas en el marco de las nuevas tecnologías de la información y la comunicación.

Creative Commons permite a los creadores escoger entre varias opciones de licencia con el fin de hacer accesible su trabajo de forma generosa. La selección de opciones se realiza mediante diferentes iconos de fácil interpretación para usuarios y ordenadores, ya que éstas vienen enmarcadas en distintas escenas de metadatos.

Creative Commons reconoce que, si bien en el mundo de la creación y la cultura, su trabajo ha sido fructífero, en el caso de la ciencia la cosa ha resultado más compleja, por tratarse de un modelo de comunicación científica regulada por un mercado de información, unos mecanismos de acreditación mediante *peer review*, protección de patentes, acuerdos de transferencia tecnológica (*Materials Transfer Agreement*), bases de datos, laboratorios virtuales, agencias de financiación, etc^{vii}.

Science Commons comenzó trabajando en tres áreas: la comunicación científica, las políticas de licencia y la web semántica. En el caso de la comunicación científica tienen en cuenta tres niveles de información: los datos generados en la investigación experimental, las publicaciones *peer review* que explican e interpretan los datos, y los metadatos que describen e interpretan los artículos relacionados con los datos. Para ello han desarrollado el proyecto **Science Commons Publishing Project**: se trata de promover el acceso abierto, facilitar los mecanismos de autoarchivo de artículos y datos (acuerdos de licencias con *PLOS, BioMed Central, Springer OpenChoice*, etc); fomento de metadatos estandarizados para las diferentes disciplinas, etc.

Otras líneas de trabajo a destacar:

-Biological Materials Transfer Project: se trata de contrarrestar las limitaciones en el acceso de la actual normativa vigente en materia de transferencia tecnológica (*Materials Transfer Agreement*) a modo de proyecto piloto en el ámbito de la biología molecular. Centrado en la reducción de tiempos y costes en los contratos de transferencia entre instituciones sin ánimo de lucro con agencias de financiación privada.

A través de una herramienta Web, los proyectos pueden absorber metadatos que posibiliten la localización, reconocimiento, así como posibles acciones futuras en su transmisión entre diferentes comunidades.

-**Neurocommons**: se trata de un proyecto en el que se auna el acceso a los materiales científicos con el software de código abierto que hace posible su transmisión y explotación. **Neurocommons** entra dentro de los desarrollos de ciberinfraestructura y *Open Knowledge*, permitiendo la explotación, emulación y simulación de datos y resultados de su plataforma. **Neurocommons** hace uso del *Open Access Protocol*, herramienta asociada a los datos científicos que permiten su integración en otras bases de datos de forma legal. Ya no se trata de una licencia, sino de un conjunto de buenas prácticas que permiten ir marcando de forma acumulativa los datos que entrarían a formar parte del dominio público en el futuro.

1.1.3. 'Principios Pantón', hacia los datos en abierto^{viii}

Se trata de una tabla o conjunto de principios presentados por Peter Murray de la Universidad de Cambridge, Cameron Neylon y Rufus Pollock de la *Open Knowledge Foundation*, y John Wilbanks de *Science Commons*, entre otros, a fin de establecer una serie de recomendaciones que posibiliten el acceso abierto a los datos científicos. Su objetivo es la difusión de los principios de la *Open Knowledge Foundation* y la implementación en la comunidades de investigación del *Open Data Protocol*. Reclaman, al tiempo, que la ciencia se traslade al dominio público, mediante la utilización de licencias **Creative Commons**, o similares.

Los 'Principios Pantón' transfieren el concepto de *abierto*^{ix} en red a los datos científicos, permitiendo a cualquier usuario su descarga, copia, análisis, reutilización en fines no previstos, sin barreras legales, técnicas o económicas. Los límites están en la integridad de los mismos y el reconocimiento de la autoría.

2.1 Recomendaciones y un posible modelo de gestión de datos para las instituciones académicas^x

El desarrollo de las tecnologías de la información y la comunicación está transformando la naturaleza y escala de los modelos de investigación por el modo operativo en que generan y consumen información, especialmente datos. La facilidad con que estos pueden ser almacenados y difundidos para otros a través de las nuevas estructuras de la ciencia significa que las instituciones deben trabajar en la línea de dar acceso, compartir e intercambiar información con el fin de aumentar su impacto y visibilidad.

Razón por la que las universidades y comunidades de investigación, la comunidad científica en su conjunto (sin olvidar, bibliotecas y archivos, con su preceptiva información a usuarios), deben diseñar políticas, consensuadas, de distribución de sus propios recursos de información, derivados de investigación con fondos públicos, de los que se ha garantizado preservación y futuros usos, además de elaborar un protocolo de buenas prácticas para la comunidad científica y académica, con instrucciones muy precisas y programas que tengan en cuenta y estén sensibilizados con estrategias de acceso, integridad de los datos, transparencia de procesos, responsabilidad profesional, interoperabilidad, protección de la propiedad intelectual y, finalmente, entendiendo siempre que la gestión debe entrar dentro de la preservación en el tiempo, problema central de la e-ciencia.

2.1.1 Algunas consideraciones.

Indicamos, a continuación, algunas ideas sobre, o pasos a dar para, este modelo de gestión de datos:

-Es importante, en primer lugar, proceder a un vaciado de instrumentos y a un análisis de las tendencias, buenas prácticas y políticas ya en curso, porque nos va a permitir construir o diseñar una estrategia e infraestructura de información robusta en contenidos y servicios con el objetivo de dar apoyo a la educación y nuevas líneas de investigación. En este sentido se han producido importantes avances en los últimos años que permiten crear una cartografía de dichos archivos de datos y sus diferentes políticas (NSB, 2005)^{xi}.

-Luego que estos, no son puras estanterías o almacenes, sino modelos de gestión de datos. Hay que conocer, por lo tanto, las diferentes actitudes, modelos y prácticas de las de comunidad investigadora, teniendo en cuenta que los procesos y propósitos de investigación varían según la disciplina y así sus metadatos, nivel de estandarización, cultura de acceso, etc. Aunque habría que trabajar con todos ellos bajo la filosofía de bien público.

-Por otro lado, en tercer lugar, a nivel organizativo, es preciso romper las barreras institucionales entre bibliotecas, archivos de datos e instituciones de investigación; asumiendo nuevos roles y reconociendo que ninguna institución por sí misma puede ser responsable de la gestión. Se tiene que trabajar a nivel local, nacional e internacional con el fin de maximizar el valor de los datos y resultados científicos de hoy y del futuro. La e-ciencia y las ciberinfraestructuras en las que esta se apoyan, sólo contempla la gestión a un nivel multinstitucional (Walters; Skinner, 2011).

-Ya a nivel de datos digitales hay que trabajar con su propio ciclo de vida, con el objetivo de asegurar su preservación en el futuro. En base a dos principios, presentes en todo el proceso: a) la ciencia debe estar basada en la apertura y libre intercambio de ideas, informes, opiniones, conocimientos y datos, toda investigación financiada con fondos públicos deben estar accesible para su uso (aquellos resultados derivados de financiación privada y sin una aplicación comercial deberían ser almacenados para posibles usos alternativos); b) bajo criterios de selección y acordes a los diferentes niveles de colección de la NSF, tanto los datos base como la información derivada deben ser preservados en el tiempo.

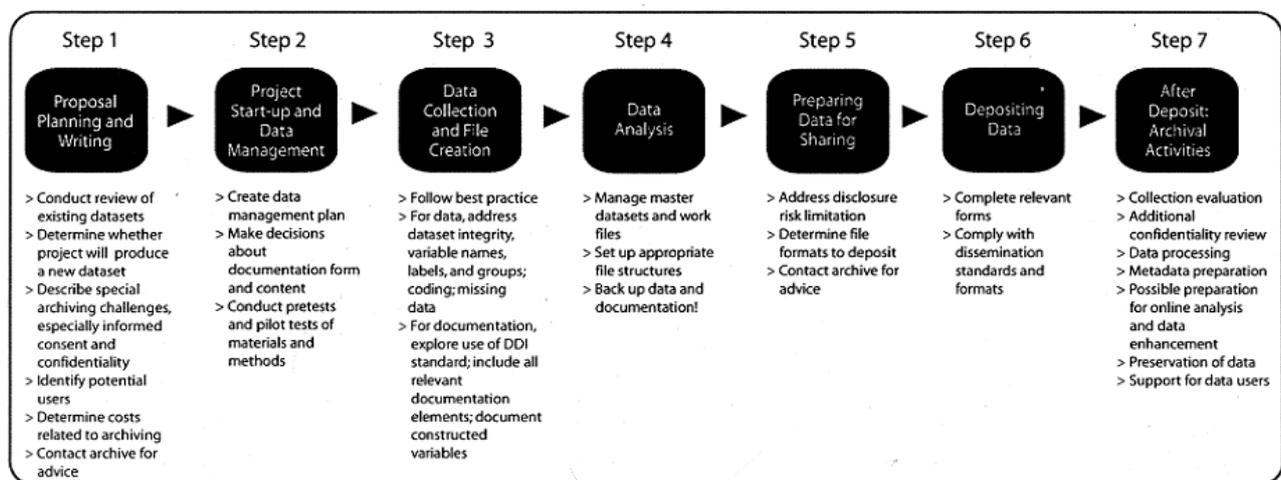


Figura 1. **Inter-university Consortium for Political and Social Research (ICPSR) The Guide to Social Science Data Preparation and Archiving (2009): The data life cycle.**

-Todo el proceso requiere de una financiación adecuada por parte de las instituciones implicadas. Los costes de la preservación deberían ir identificados de forma específica en la financiación de cualquier proyecto de investigación siguiendo las pautas del *Blue Ribbon Task Force*. Habrá que incluir siempre una partida, pues, específica destinada a la gestión de datos.

-Bibliotecas, Centros de datos e instituciones vinculadas a la investigación deben proporcionar una formación clara a los involucrados en el proceso mediante alfabetización en información científica, edición de guías y manuales, creación de plantilla, etc.^{xii}

-La clave está en el propio ciclo de vida de la información: las partes interesadas deben trabajar, en este sentido, con modelos concisos y estandarizados a lo largo de todo el proceso o recorrido de la investigación, desde la planificación a la difusión (creación de datos, análisis, resultados, conservación, prestación de servicios, etc.) Todo plan de gestión de datos debe incidir básicamente en dos puntos estratégicos: en la creación del proyecto de investigación y en la difusión de los resultados del mismo.

-Dos coordenadas, la una instrumental y la otra social, debieran marcar este programa de gestión: a) la aceptación de *Reference Model for an Open Archival Information System (OAIS)*; b) la consideración de la información en todo momento como bien público. Dos criterios muy elementales, pero irrenunciables.

Conclusiones

La e-ciencia es un universo infinito y preñado de posibilidades de enriquecimiento informacional. En nuestras manos está el incrementar o empobrecer su poder operacional y de transformación social.

Las dos claves para aprovechar estos recursos de una forma intensiva y eficaz serán: una planificación inteligente en la gestión de datos y el incremento de la cooperación internacional en la difusión e intercambio de los mismos en base a un conjunto de instrumentos que favorezcan coordinación, uniformidad y agilidad.

El primer punto, crear una plataforma adecuada para que datos y resultados de la investigación puedan estar disponibles a la comunidad científica para su reutilización e intercambio, hace preciso que estos sean adecuadamente tratados y almacenados, de forma preferente en los primeras etapas de un proyecto de investigación (idealmente ya en la planificación inicial, como ya hemos subrayado anteriormente), identificando usos, licencias, posibles archivos de almacenamiento, etc. De esta manera se incrementaría el valor, la visibilidad, el impacto, su disponibilidad en el tiempo y el acceso a los mismos.

El segundo punto, la cooperación internacional entre instituciones, está en proceso de reinención, pero ya contempla toda una serie de instrumentos estandarizados y de amplio consenso en la comunidad científica. Nos referimos aquí al tipo de políticas como las empleadas por varias organizaciones internacionales: caso de la *Data Seal of Approval* by the *Data Archiving and Networked Services (DANS)* en los Países Bajos, los niveles de colección establecidos por la *National Science Foundation (NSF)* y, en el orden de las buenas prácticas, no debemos olvidar nunca las recomendaciones de sostenibilidad de la *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* y procurar adherirse a la ruta de la *Open Knowledge Foundation*.

Lo cierto es que, cada institución, disciplina y proyecto es un mundo propio y original, pero de lo que se trata es de convertir las diferencias en oportunidad de intensificar y potenciar un canon común de procesos. El reto está ahí y resulta ineludible, ya que nos encontramos ante un cambio radical de paradigma y una mutación epistemológica sin parangón hasta el momento. En nuestras manos está orientarlo en un sentido u otro.

Referencias bibliográficas:

Blue Ribbon Task Force on sustainable digital preservation and Access/ National Science Foundation; The Andrew W. Mellon Foundation; Library of Congress; Joint Information Systems Committee; Council on Library and Information Resources; National Archives and Records Administration (2008). Sustaining the digital investment: issues and challenges of economically sustainable digital preservation. Consultado en: 1-Sept. – 2011. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

Blue Ribbon Task Force on sustainable digital preservation and Access/ National Science Foundation; The Andrew W. Mellon Foundation; Library of Congress; Joint Information Systems Committee; Council on Library and Information Resources; National Archives and Records Administration (2010). Sustainable economics for a digital planet: Ensuring long-term Access for digital information. Consultado en: 1-Sept. – 2011. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.

CODATA. *Bits of Power: issues in global access to scientific data.* Washington, D.C.: National Academy Press, 1997, ISBN: 10: 0-309-05635-7.

Inter-university Consortium for Political and Social Research, ICPSR (2009). *The Guide to Social Science Data Preparation and Archiving: best practice through the data life cycle.* 4th ed. Consultado en: 1-Sept. – 2011. <http://www.utwente.nl/igs/DataLab/Documents/dataprep.pdf>

Jenny Fry, Suzanne Lockyer, Charles Oppenheim, John Houghton and Bruce Rasmussen. Centre for Strategic Economic Studies, Victoria University, Melbourne; Department of Information Science, Loughborough University (2008). *Identifying benefits arising from the curation and open sharing of*

- research data produced by UK Higher Education and research institutes. Consultado en: 1-Sept. – 2011. http://ie-repository.jisc.ac.uk/279/2/JISC_data_sharing_finalreport.pdf
- Joint Information System Committee (2010).** *Keeping Research Data Safe*. Consultado en: 1-Sept. – 2011. <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>
- Joint Information System Committee (2010).** *Managing and sharing data: a best practice guide for researchers* / Veerle Van den Eynden, Louise Corti, Matthew Woollard and Libby Bishop. Consultado en: 1-Sept. – 2011 <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>
- League of European Research Universities, LERU (2011).** *The LERU Roadmap toward Open Access. Advice Paper, n° 8*. Consultado en: 1-Sept. – 2011. http://www.leru.org/files/publications/LERU_AP8_Open_Access.pdf
- National Institutes of Health (2003).** *NIH Data Sharing Policy and Implementation Guidance*. Consultado en: 1-Sept. – 2011. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
- National Science Foundation (2011).** *Changing the Conduct of Science in the Information Age*. Consultado en: 1-Sept. – 2011 <http://www.nsf.gov/pubs/2011/oise11003/>
- National Science Board, NSB (2005).** *Cyberinfrastructure Vision for 21st Century*. Consultado en: 1-Sept. – 2011. http://www.arl.org/bm~doc/ci_vision_march07.pdf
- National Science Foundation (2005).** *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*. Consultado en: 1-Sept. – 2011. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540_10.pdf
- Organisation for Economic Co-operation and Development (2007).** *OECD principles and guidelines for access to research data from public funding*. Consultado en: 1-Sept. – 2011. www.oecd.org/dataoecd/9/61/38500813.pdf.
- Research Information Network (2010).** *Open to all? : case studies of openness in research*. Consultado en: 1-Sept. – 2011 http://www.apo.org.au/sites/default/files/NESTA-RIN_Open_Science_V01.pdf
- Research Information Network (2008).** *To Share or not to share publication and quality Assurance of Research Data Outputs. Report to the Research Information Network* / Swan, A and Brown, S. Consultado en: 1-Sept. – 2011. <http://eprints.ecs.soton.ac.uk/16742/>
- Sell, Susan K.** «Intellectual Property Rights». En: Held, David; McGrew, Anthony. *Governing globalization: Power, authority and global governance*. Cambridge, UK : Polity Press, 2007, pp. [171]-[188]. ISBN [978-0-7456-2733-5].
- UKADATA, UK Data Archive (2011).** *Managing and Sharing Data: a best practice guide for researchers*. Consultado en: 1-Sept. – 2011. <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>
- Varela Orol, C.** «Acceso libre ó coñecemento?: políticas neoliberais nas bibliotecas universitarias galegas». En: *A Trabe de Ouro* , 2005, n. 64, pp. 481-496.
- Walters, Tyler; Skinner, Katherine.** *New roles for new times: digital curation for preservation*. Washington, D.C.: Association of Research Libraries, 2011, ISBN: 1-59407-862-9.
- Wellcome Trust (2010).** *Wellcome Trust policy on data management and sharing*. Consultado en: 1-Sept. – 2011. www.wellcome.ac.uk/Aboutus/Policy/Policy-and-position-statements/WTX035043.htm.
- Wilbanks, John; Boyle, James (2007).** *Introduction to Science Commons*. Consultado en: 1-Sept. – 2011. http://sciencecommons.org/wp-content/uploads/ScienceCommons_Concept_Paper.pdf

ⁱ Tres son los niveles funcionales o tipologías de colección que presenta la *NSF* (estos niveles son elemento y premisa para cualquier modelo de selección en un archivo digital sostenible): 1^a) «Colecciones de datos de investigación», o *Research data collections*: los resultados de una investigación. Generalmente, son objeto de una preservación limitada en el tiempo, siendo las normas y formatos comunes en una comunidad dada, pudiendo o no adaptarse a los estándares y metadatos de la misma. A menudo, al ser una información novedosa, no existen sustitutos. Puede que no haya necesidad de preservar la misma más allá del proyecto de investigación. Estas colecciones por lo general cuentan con presupuestos de financiación bajos. La *NSF* y *JISC* (*Joint Information System Committee*, 2010) recomiendan políticas de selección y retención por parte de los autores, en algunos casos con medidas legales; 2^a) «Colecciones de datos base de una comunidad» o *Resource or community data collections*: colecciones vinculadas a una disciplina. La propia comunidad científica establece ciertos parámetros, normas y estándares para su creación, acceso y uso. Los presupuestos vinculados a estas colecciones son de nivel intermedio y financiado a través de agencias gubernamentales. Debido al cambio en la financiación del proyecto y la vida de este, es difícil en ocasiones establecer las políticas de retención. La *NSF* y el *JISC* recomiendan la colaboración entre autores e instituciones a fin de establecer acciones en este sentido y dirección; 3^a) «Colecciones de datos referenciales» o *Reference data collections*: colecciones fundamentales para una comunidad científica. Están vinculadas a un número de usuarios amplio, más allá de una institución y generalmente de alcance internacional. Las normas de creación son rígidas y con gran nivel de especialización. Los presupuestos de financiación son elevados y provienen de diferentes fuentes. La preservación de estas colecciones recae en diferentes organismos, no siempre estables. Estas colecciones pese a ser de referencia para una comunidad dada, son las que presentan mayores barreras en cuanto a su acceso

ⁱⁱ Algunos ejemplos: los *National Institutes of Health, NIH*, en su *Data Sharing Policy* (2003), consideran que el acceso a los datos científicos es esencial para traducir la investigación en conocimiento, productos y procedimientos que lleven a la mejora de la salud; la *NSF*, incide en esa necesidad de poder acceder y compartir los datos base («raw data»), así como los resultados de las investigaciones en curso que hayan sido financiadas con fondos públicos en un tiempo razonable y con un coste sostenible. (En: *National Science Board, National Science Foundation*. (2005). «Current Policies on Data Sharing and Archiving». *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*. Pp.: 57-71). Otros ejemplos de retención de los resultados científicos se pueden ver en *Wellcome Trust* (2010): *Wellcome Trust policy on data management and sharing*; y en *Organisation for Economic Co-operation and Development* (2007): *OECD principles and guidelines for access to research data from public funding*.

ⁱⁱⁱ Tomemos, por ejemplo, la línea de argumentación de *UKDATA, UK Data Archive: Sharing your data – Why and how*, en: *Managing and Sharing Data: a best practice guide for researchers*: «Los datos científicos son un recurso valioso, por lo general requieren para su producción grandes cantidades de tiempo y dinero. Muchos de los datos tienen un valor significativo más allá del uso programado en la investigación original. Compartir/intercambiar los datos en la investigación: a) fomenta la investigación científica y el debate; b) promueve la innovación y nuevos usos en los mismos; c) fomenta nuevas colaboraciones entre las comunidades científicas; d) maximiza la transparencia y la rendición de cuentas; e) permite el escrutinio de los resultados de la investigación; f) mejora y fomenta la validación de la investigación y sus métodos; g) aumenta el impacto y la visibilidad de la ciencia; g) promueve líneas de investigación; h) puede proporcionar nuevos financiamientos; i) ofrece importantes recursos para la educación y la formación».

^{iv} El movimiento *Open Access* reclama la necesidad de dar acceso a los resultados de la investigación científica en internet, de tal manera que se puedan leer, descargar, copiar, distribuir, imprimir, buscar, *text mined* o bien usarlos para otros fines educativos, culturales y científicos sin ninguna barrera económica, legal o técnica. Desde la *Declaración de Budapest* (2002), el movimiento de acceso en abierto se ha ido extendiendo a nuevos ámbitos. En la actualidad, se empieza a hacer la distinción entre *Open Scholarship* (acceso a los resultados de una investigación) y el *Open Knowledge* (concepto más amplio que englobaría cualquier tipo de información relativa a los datos y resultados de la e-ciencia: de esta manera se trabaja en el libre acceso a los datos primarios (*raw data*), el *software* asociado y la interpretación de los mismos. En: *Open Knowledge Foundation: <http://okfn.org/>*).

^v «Propiedad intelectual en relación a los Datos: ¿restricciones legales en relación al acceso abierto? La aparición de un nuevo modelo de propiedad intelectual que protege de manera restrictiva el contenido de las bases de datos, así como su publicación está afectando de manera significativa a la circulación de los datos científicos en la comunidad internacional. El problema se está debatiendo a nivel internacional, a fin de establecer un marco jurídico que no subordine los intereses del bien público a [ante] los que buscan la protección de la explotación de estos. Desafortunadamente, hasta hace poco, el avance legislativo a favor de la educación y el avance científico ha sido casi inexistente. Es necesario que se lleve a cabo una labor sostenida por parte de los sectores implicados que eviten las posibles restricciones en el acceso abierto y el intercambio pleno de la información científica». En : *Bits of Power: issues in global access to scientific data / U.S. National*

Committee for CODATA. Washington, D.C.: National Academy Press, 1997. P.: 8 (En la misma monografía: *Recomendaciones en relación a las propuestas legales que afectan a el acceso a los datos científicos*. P. 15).

^{vi} Modelo, en parte, para la *National Science Foundation (2005)*. *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century* y los *OECD principles and guidelines for access to research data from public funding*.

^{vii} «Tenemos la ambición de enfrentarnos al mundo de la ciencia y los datos, de la misma manera que *Creative Commons* ha trabajado en el mundo de la cultura y la educación: erosionando barreras legales y técnicas innecesarias al acceso, promoviendo la innovación, dando herramientas fáciles, sencillas y de calidad para individuos e instituciones mediante el uso de términos específicos acorde al grado de deseo de compartir de los mismos». En: Wilbanks, John (2006). *Introduction to Science Commons*. P.6

^{viii} En: <http://pantonprinciples.org/>

^{ix} Se considera una obra en abierto a aquella que permite en su distribución, entre otras acciones: el acceso íntegro; la redistribución; reutilización con otros propósitos (manteniendo la integridad y el reconocimiento de autoría); la ausencia de restricciones tecnológicas y la no discriminación a personas o grupos de personas.

^x Este modelo de gestión está basado fundamentalmente en los patrones y guías del *Joint Information System Committee (2010)*; *Inter-university Consortium for Political and Social Research, ICPSR (2009)* y en el examen del *Research Information Network (2008)*.

^{xi} Apéndice E. *Representative Distributed Research Communities (Virtual Organizations)*. NSB, *National Science Board (2005)* *Cyberinfrastructure vision for 21st Century*. P. 50-56

^{xii} Se pueden ver ejemplos de formación en las Bibliotecas del MIT: *Data finding and managing*: <http://libguides.mit.edu/finding-data/> ; así como en los programas presentados conjuntamente en el Reino Unido por organizaciones como el *Digital Curation Centre (DCC)*, el *UK Data Archive (UKDA)* y el *JISC : Research data management training materials (RDMTrain)*; donde podemos encontrar herramientas de formación y pautas a seguir para nuestras instituciones: http://www.jisc.ac.uk/whatwedo/programmes/~link.aspx?_id=677B9C0D0E8F4B12A7E2ACC86FD9D736&Z=Z