

LA RED PERMANENTE: CIENCIA EN INTERNET

Giorgio de Marchis

La Red Permanente: Ciencia en Internet

Giorgio De Marchis

(Ruego se me envíe la referencia de todo artículo-libro-ponencia en el que se vaya a citar este artículo. Muchas gracias.

Se puede mandar a: giorgio_de_marchis@yahoo.com ó a: giorgiodemarchis@hotmail.com)

Palabras clave:

Ciencia; Citas; Conservación; Depósito; Digital; Enlace; Enlace roto; Hemeroteca; Identificador; Internet; Permanencia.

Key words for search in the Internet:

Cite; Citation; Deposit; Digital; Broken links; Identificator; Internet; Linkrot; Links; Permanence; Preservation; Science; Repository.

Tema: Accesibilidad en el largo plazo de los materiales publicados en Internet. Necesidad de una solución para que pueda evolucionar la Ciencia en Internet.

Resumen: El problema que se trata en este trabajo es el siguiente: los investigadores no pueden citar artículos u otra información publicada en Internet si no pueden estar seguros que en el futuro la información seguirá estando disponible. Se analiza el estado de la cuestión: este problema no se ha tratado aún en España y se ha tratado poco de una forma explícita en investigaciones internacionales. La mayoría de las discusiones entabladas hasta la fecha se han centrado en los enlaces rotos y la necesidad de dar un “DNI” a cada documento; los derechos de autor; la necesidad de las Bibliotecas de conservar los documentos. Se hace una recopilación de lo que se está haciendo actualmente para dar una solución a algunas de estas cuestiones. Terminamos la exposición con una serie de propuestas de diferentes autores y una propuesta propia para la solución del problema planteado.

Un ejemplo de Umberto Eco

Supongamos que estás leyendo “Opera aperta” (Eco 1988). Supongamos que has llegado a la página 35, y encuentras la referencia a un artículo de Pousseur. Eco, como todo buen investigador, ofrece a pie de página la referencia: *La nuova sensibilità musicale*, in “Incontri musicali”, n.2, maggio 1958, pag. 25. Supongamos además que estás preparando un artículo sobre la sensibilidad musical en los niños de 6 a 10 años. Previsiblemente desees leer el texto original. Eco te ha dado el nombre del autor; el título del artículo y de la revista en la que se encuentra; el número (en este caso no existe volumen); así como el año de publicación y la página en el que se encuentra el artículo. Ahora tienes que conseguir el artículo: es el problema de todo investigador. Tratándose de un artículo en una revista italiana de los años 50, parece ser que la mejor solución desde España sea la de conseguirlo gracias a la ayuda de los siempre tan amables y eficientes bibliotecarios. Si no se consigue a través de la biblioteca de la Facultad, probarías a través de la Hemeroteca Nacional. A lo mejor tendrás que pagar por una copia del artículo. Gracias a la eficiencia de los bibliotecarios y del servicio de Préstamo Interbibliotecario, lees el artículo. Posiblemente haya pasado un mes desde que lo has buscado.

Este proceso lo han seguido los investigadores durante décadas. Con frecuencia, ahora es mucho más rápido. Las bases de datos no relacionales han proliferado en la Red de Redes. A la hora de hacer la cita, un hipervínculo te remite a la bibliografía o, como es deseable, te debería remitir directamente al texto completo. Si esto no ocurre, la mayoría de las revistas científicas tienen su base de datos en Internet, y puedes acudir a ellas directamente. Además, están surgiendo diferentes proyectos que promueven el acceso libre a los artículos científicos a través de la Red de Redes.

Planteamiento del problema

Muchos piensan que las revistas electrónicas son el futuro de la investigación científica. Las publicaciones de muchas revistas internacionales se venden (y guardan) a través de bases de datos gestionadas por empresas privadas. Los particulares, y sobre todo las instituciones (Universidades, centros de investigación, etc.) se suscriben a ellas. La Universidad para la que investigas paga todos los meses por unas suscripciones a bases de datos virtuales que te permiten encontrar y descargarte los artículos inmediatamente. Los trabajos que se encuentran en estas bases de datos posiblemente se vayan a poder recuperar dentro de 50 años, tal y como ocurría con el artículo de Pousseur¹.

¿Qué ocurre sin embargo con todos aquellos artículos que están en revistas electrónicas de menor importancia? ¿Con las revistas electrónicas no científicas? ¿Y qué pasa con aquellos artículos publicados que no son propiamente científicos? Pongo como ejemplo un artículo escrito para una

¹ Puede ser que estas empresas quiebren, sin embargo es difícil que su Base de Datos vaya a perderse, siendo lo más probable que otra empresa del sector la compre y siga gestionándola.

revista editada en Internet y que explica el funcionamiento de un determinado disco duro.

Supongamos que alguien escribe un artículo acerca del citado disco duro. Supongamos que este artículo es leído dentro de un año por un investigador que utiliza la comodidad de las búsquedas en la red para escribir sus trabajos científicos. Ha entrado en la revista, ha encontrado el artículo. Necesitaba información² acerca del funcionamiento de aquel disco duro. Copia las frases y los datos del artículo y lo cita en el suyo. Los cita tal y como aconseja la APA para las publicaciones recogidas de Internet. Supongamos que el artículo de nuestro investigador se publica en una revista impresa al ser todavía las únicas publicaciones a las que el mundo académico otorga valor.

Dentro de 50 años, otro investigador, tal y como nos ha ocurrido con el libro de Eco, encuentra la cita y piensa que se puede tratar de un artículo interesante para una investigación que está realizando acerca del desarrollo de los discos duros. Pero cuál es su sorpresa cuando ve que la página web ya no existe...y que no existe ningún lugar en el que pueda recuperar el artículo.

Este es el problema que queremos apuntar en el presente artículo: no hay nada que asegure a un investigador que las citas que hace de los artículos publicados en Internet, se puedan volver a encontrar dentro de 50 años, ni siquiera se puede asegurar que vayan a poder encontrarse dentro de un año. Esta situación genera incertidumbre, que a su vez provoca que los investigadores eviten citar artículos y publicaciones de Internet. Es necesario crear un sistema que asegure la permanencia de los documentos publicados en la Red de Redes.

Llamaremos a este problema como *el problema de la permanencia*.

Importancia del problema

Dentro del método científico tiene gran importancia el concepto de “replicabilidad”, “es decir, que (toda investigación) pueda ser sometida a revisión e incluso reproducción en su planteamiento, desarrollo y ejecución” (Sierra Bravo). La elaboración de las referencias, es uno de los elementos que caracteriza a los trabajos científicos, y que tiene como una de sus principales funciones la replicabilidad³. Si no se puede cumplir con esta condición, el valor científico de una investigación es menor.

Para preparar este artículo he estado buscando ensayos en la red. A través de un buscador he encontrado uno con el título: *Internet en la investigación científica*. Pero al abrir el enlace, aparece la conocida página con escrito: “No se puede

² Y en otros casos, unas reflexiones, que por supuesto no son exclusivas de las personas que nos dedicamos a investigar.

³ Señala Sierra Bravo (Ibidem) que otras funciones de las citas pueden ser dar a conocer las fuentes primarias de la investigación; expresar las ideas de otros autores que, muchas veces gracias a su “autoridad” pueden servir de apoyo a las nuestras o de contrapeso; también sirven simplemente para conocer el estado de la situación y cuáles son los trabajos no consultados.

mostrar la página. La página Web solicitada no está disponible en este momento. Puede que el sitio Web tenga problemas técnicos o que necesite ajustar la configuración de su explorador”. Dos meses después sigue sin aparecer. Lo que realmente ha ocurrido es que habían cambiado la dirección de la página o la habían suprimido.

David J. Salomon (Salomon 1999) comenta: “Con el constante cambio de formatos digitales y la rápida proliferación y muerte de las revistas electrónicas, importante material científico podría no ser preservado”, y posteriormente dice que “los investigadores son renuentes a publicar sus mejores investigaciones electrónicamente por la preocupación de no obtener el crédito apropiado. Al mismo tiempo la comunidad de investigación no concederá crédito académico significativo a las publicaciones electrónicas hasta que perciba que se está publicando material de alta calidad en revistas puramente electrónicas. A medida que un número significativo de manuscritos de alta calidad comiencen a aparecer en revistas puramente electrónicas, esta percepción empezará a terminarse”.

Dídac Martínez (Martínez 2003) habla de los altos costes que tienen las Universidades a la hora de suscribirse a las revistas y bases de datos científicas. Con la llegada de las revistas electrónicas se esperó que el panorama fuera a cambiar. Las ventajas de una revista electrónica son muchas⁴, así que se pensó en un primer momento que los precios de las publicaciones bajarían por los reducidos costes de publicar una revista electrónica, pero no es así; los precios de las revistas “consolidadas” siguen subiendo. Como alternativa a esto, diferentes grupos de investigadores y Universidades están creando revistas electrónicas de acceso libre o a bajo coste. La insatisfacción por la situación actual, las posibilidades que ofrece internet y la buena acogida que están recibiendo estas nuevas revistas “abiertas” parece hacer suponer que habrá cambios en el medio plazo. Parece probable que muchas revistas impresas desaparecerán y nacerán algunas nuevas, sólo digitales. (Polanco Masa 2002) “No cabe duda, en el futuro casi la totalidad de las publicaciones científicas se nos mostrarán en versión electrónica exclusivamente. Lo que no sabemos es cuán caótica será la transición desde las versiones impresas hasta el imperio de las publicaciones virtuales”.

Es necesario solventar el problema de la permanencia para liberar los artículos de la publicación impresa. Hasta que esto ocurra, la ciencia no podrá asociarse a la red. Se trata entonces de un círculo vicioso que se está resolviendo poco a

⁴ Dídac Martínez cita algunas: “La edición digital de las revistas científicas aporta tantas ventajas al investigador que pronto pasarán a la historia las revistas en soporte papel. De hecho, empiezan a no editarse. Las revistas electrónicas aportan una serie de soluciones para la misma labor del investigador que no puede ofrecer la versión en papel. Mayor rapidez en todos los niveles, acceso y búsqueda de un determinado artículo con potentes ‘metabuscadores’, rapidez en el envío de un artículo a la editorial, posibilidad de manipular, tratar y trabajar con los contenidos, recibir sólo aquello que interesa en el propio buzón electrónico, intercambiar opiniones con otros investigadores en tiempo real, acceder a sumarios electrónicos y grandes archivos de *pre-prints*, artículos interconectados con otras revistas o bases de datos incluso con los mismos autores o centros de investigadores y un largo sinfín de ventajas que aún serán más perfeccionadas en el tiempo inmediato”.

poco gracias a los nuevos planteamientos de los que voy a hacerme eco, después de acotar de una forma más precisa nuestro problema.

Acotamiento del problema

Vamos a limitarnos a abordar el problema de una forma general y programática. Nos centramos sólo en las publicaciones con intención de transmitir información solvente⁵ (especialmente artículos) que se publican digitalmente por Internet. Esta aproximación quiere dejar fuera muchos grandes campos de discusión, como por ejemplo el Net-art, y muchos problemas relacionados con la tecnología y los derechos de copyright⁶.

El problema de la *desinformación*

Alguien ha dicho que las buenas páginas web no se extinguen. Esta es una concepción darwinista de la web, que ha heredado también el sistema capitalista. No sólo puede haber páginas con buenos artículos pero poco conocidas que desaparecen, sino que las páginas así llamadas “buenas”, pueden cambiar de estructura y llegar a desaparecer. Sabemos que muchas revistas de gran calidad o importancia histórica mueren. El cierre de una revista puede ser debida a muchos factores ajenos a la calidad de la misma: cualidades del editor, buena relación entre los socios, modas, etc.

Cuando se habla de que en la Red hay mucha información de baja calidad⁷, sin contrastar y que sólo perjudica a la investigación seria, y que entonces es bueno que desaparezca, se está tocando un problema muy real. Sin embargo esta cuestión (que también tiene que encontrar soluciones), es un problema distinto del que estamos discutiendo hoy. El hecho de que muchos de los artículos que se protegerían si se solucionara el problema de la permanencia serían de nula calidad, no nos atañe aquí. Lo importante es que tenemos que tener un sistema para proteger las referencias de los trabajos publicados en la red, y no sólo de los trabajos publicados en las revistas de reconocido prestigio.

Manera de citar

La cuestión de cómo citar en Internet es algo que también es colindante con nuestro tema. Sobre cómo citar los documentos publicados en Internet, la

⁵ Especialmente artículos, pero también comunicaciones y conferencias, informes técnicos y libros.

⁶ Por ejemplo: ¿Qué estrategias de conservación de archivos digitales se adoptarían? ¿Cómo se actualizarían los soportes? ¿En qué medida sería aceptable modificar las fuentes en cuanto fuera imprescindible hacerlo para poder archivarlas? I.e. transformarlas en XML. Ver. M Bide, E. J. P., A Watkinson (1999). Digital Preservation. An introduction to the standards issues surrounding the deposit of non-print publications, The Library and Information Commission: 45.

⁷ Polanco Masa (2001): “Todavía pocas revistas, con formato digital exclusivo, han alcanzado prestigio. Esto va cambiando rápidamente pero hay peligros surgidos, precisamente, de lo fácil y económica que es este tipo de edición. Se ha llegado en la red a la cultura de la permisividad total. Todo el mundo puede publicar lo que quiera. Encontrar informaciones serias entre tanta bazofia será una tarea cada vez más difícil”.

fuente de mayor valor es el [ISO 690-2](#). Sin embargo, al decir en la misma página que: “The data recorded in the reference shall refer to the specific copy of the document that was seen or used. For example, the data in a bibliographic reference to a document on the Internet shall reflect the title, dates, location information, etc. for the particular version that was seen and the particular network location from which that version was accessed”, está creando una confusión que no debería existir. A saber, si citamos un artículo en la página en el que lo encontramos, es muy posible que dentro de algunos años haya desaparecido, no tenemos referencia segura.

Más tarde se habla de la forma de citar la “availability and access” y se dice:

“Information for identifying and locating the source of the document cited shall be provided for online documents. This information should be identified by the words ‘Available from’ or an equivalent phrase. The location information for online documents within a computer network such as the Internet shall refer to the copy of the document that was actually seen and should include the method of access to the document (e.g. FTP) as well as the network address for its location”.

Durante bastante tiempo, y aún ahora es frecuente, se han citado los trabajos en Internet de la siguiente forma: *Electronic journal articles* Loker WM (1996) "Campesinos" and the crisis of modernization in Latin America. *Jour Pol Ecol* 3(1). Available: http://www.library.arizona.edu/ej/jpe/volume_3/asciilokeriso.txt via the Internet. Accessed 1996 Aug 11.

Este tipo de citación equivale a decir: “si tu buscas el artículo en el 2005 y no lo encuentras, lo siento pero no es mi culpa: en 1996 estaba allí”. A todas luces este planteamiento es insuficiente.

No parece que se hayan dado cuenta del problema que estamos planteando en este artículo.

Intento de registrar todo lo que existe en la web

Poca relación con lo que comentamos tiene la base de datos [Internet Archive Building](#), que quiere recopilar todo lo que se encuentra en la Internet en un momento determinado para poder ser consultado en el futuro por los investigadores. Alejandro Polanco (Polanco Masa 2002) dice que

“esta idea se enfrenta a muchos problemas técnicos y jurídicos. Puede que en un futuro, los cientos de terabytes que almacena este ‘archivo de Internet’, sean ordenados para una consulta racional. De momento, se limita a ‘indexar’ cuantos contenidos copian sus programas robot-araña. Estos programas exploran la red capturando documentos y enlaces buscando referencias cruzadas. Cada cierto tiempo se repite el proceso, copiando de nuevo los mismos documentos, guardando los cambios y buscando nuevos enlaces. Pero la recolección del patrimonio de Internet no puede ser exhaustiva. Muchos lugares de la red, protegidos por autores y editoriales, restringen el acceso, con lo que es imposible copiar al archivo sus contenidos. El verdadero objetivo del proyecto, dados estos impedimentos, será mostrar el aspecto que ofrece Internet en un periodo determinado”.

A parte los usos para los historiadores, parece que no tiene más relación con lo que hemos planteado.

Obsolescencia de soportes

Tampoco nos vamos a interesar por el problema de mantener los datos a lo largo del tiempo debido al cambio de la tecnología que hace que, por ejemplo, encontrar un reproductor de vinilos dentro de cincuenta años, va a ser algo muy difícil, o peor aún, poder pasar los datos que están en un formato a otro: i.e. encontrar ahora una máquina que pase la información de una tarjeta de datos al disco duro de un ordenador actual (Thomas 2003).

Nuestro problema se divide en dos partes: la permanencia de los enlaces y la de los contenidos.

Veamos las soluciones que se están dando al problema de los enlaces.

Qué se ha hecho respecto a la permanencia de los enlaces

Se ha estado hablando desde hace muchos años acerca de la necesidad de tener unas direcciones permanentes. Diferentes autores y organizaciones han ofrecido varias propuestas. Algunos de estos sistemas hace referencia al conocido ISSN. El ISSN parece a todas luces insuficiente para las necesidades de información actuales y tiene que evolucionar. Así lo ha entendido el ISO y se están aprobando medidas para añadir números para que el ISSN contenga más información (metadata). Sin embargo, y porque el ISSN no resuelve el problema que planteamos, el camino parece que va más bien hacia las direcciones que adelante reseñamos.

La URL (Uniform Resource Locator) se puede comparar con una dirección postal (Payette 1998), que al igual que la residencia en la que se vive, puede cambiar. Si alguna persona te mandara una carta por ejemplo a la dirección en la que vivías de pequeño, lo más probable es que el servicio de correos la devuelva al remitente, con un sello de “destinatario desconocido”. Es lo mismo que ocurre con las páginas web que cambian de dirección. Mas, si el servicio de correo estuviera avisado que has cambiado de dirección y sabe cuál es la actual, te puede hacer llegar la carta. Esta es una filosofía similar a la que subyace en la mayoría de los métodos actualmente utilizados para obtener direcciones permanentes.

Algunas soluciones propuestas para la permanencia de los enlaces son⁸:

1. Handle System
2. PURL (Persistent URL)
3. URI - URN
4. Bibliographic Protocol (BibP) and Universal Serial Item Name (USIN)
5. Digital Object Identifier (DOI)

⁸ Para un listado amplio de herramientas creadas para identificar los objetos en Internet, ver el informe disponible en la página de mEDRA: (2002). Media-related Identification and Metadata Standards, EDRA – European DOI Registration Agency.

Algunos sistemas que no hemos visto son: OpenURL; XRI.

Handle System

El sistema se implementó por primera vez en el otoño de 1994 (Dyson 2003). Es un *resolution system*: un sistema para crear, administrar y “resolver” “unique identifiers” y que soporta los identificadores⁹. Es un sistema similar al de los PURL, pero desarrollado por la Corporation for National Research Initiatives (CNRI). Para conseguir hablar con una persona por teléfono, se busca el número en el listado de las Páginas Blancas. El número se localiza a través del nombre y apellido completo del propietario. Esto se llama “resolución”(resolution). Se trata de que el nombre corresponde a un número. Si para buscar a tu amigo de la infancia y sólo tienes el nombre de pila y el teléfono antiguo, no podrás conocer el nuevo teléfono. Básicamente se trata de dar un nombre a un objeto (el nombre y apellido) para que, aunque se cambie de teléfono (la dirección URL), se pueda encontrar el actual. Aunque alguien cambie de número de teléfono, no suele cambiar de nombre y apellido.

Los handles simplemente son unos nombres de objetos digitales. El Handle System permite encontrar a partir de un número (por ejemplo, un URN) una fuente de información, generalmente una URL. Se engloba dentro de las tecnologías conocidas como *n 2 l* (urn to url) *technologies*. El Handle System se encarga de encontrar la localización y la información asociada a los handles. Handle permite una resolución múltiple y puede dirigir a más de una fuente de información¹⁰.

El Handle System en principio se adecua al esquema de los URN teniendo un *naming scheme* para unos identificadores únicos que en este caso se llaman handles; un *resolution system* para trasladar los handles hacia la información asociada; un registro general (*registry*) que consiste en un servicio central (*Global Handle Registry* o GHR)¹¹ en el que se administran los “productores” de handles, que son los *naming authorities*. Así, todo nombre del handle tiene dos partes: 1. Un *naming authority*, también conocido como el prefijo handle (handle's prefix); es algo parecido al código de país en los números de teléfono, aunque cada “productor” de handles puede tener uno: una revista, una universidad, etc. 2. Un *unique local name*, el ID del objeto, bajo la naming authority, conocido como su sufijo (suffix) y que correspondería al verdadero número de teléfono.

Una de las tareas del Global Handle Registry (GHR) es la de asegurar que todos los naming authority sean diferentes dentro del Handle System. El GHR entrega un naming authority en el momento de la instalación.

Se pueden utilizar también los servidores proxy para trasladar handles a URLs. En este caso el Handle Proxy Server funcionaría de una forma similar al PURL resolver. CNRI y OCLC han colaborado para permitir interoperabilidad entre [Handles y PURLs](#).

⁹ Existe también el Indecs (Interoperability of Data in E-commerce Systems), que no estudiamos aquí.

¹⁰ Los DOI, que luego veremos, utilizan esta capacidad de los handles.

¹¹ El Global Handle Registry es gestionado por el CNRI, y funciona como los servicios handle locales, aunque su principal función es la de saber todo acerca de los otros servicios handle y como encontrarlos.

En la [Library of Congress](#), por ejemplo, se utilizan handles para identificar los objetos, dentro del [National Digital Library Program](#) (NDLP). _En la página web del CNRI se puede encontrar más información de [proyectos que utilizan handles](#).

PURL

El PURL (Persistent Uniform Resource Locator), ha sido desarrollado por la OCLC (Online Computer Library Center, Inc.), y está en funcionamiento desde enero de 1996. La PURL es todavía una URL a la que se le asegura permanencia. Funciona mandando la búsqueda no directamente a la URL, sino a través de una máquina intermedia que comprueba la dirección actual, y la reenvía al cliente. Se trata simplemente de redireccionar a través del comando HTTP. Para poder usar el PURL es necesario registrarse en un “resolver” (servidor de PURL), y rellenar un formulario. Se obtiene así una dirección PURL. La persona que busque la página URL irá siempre al servidor PURL que redireccionará la búsqueda a la dirección URL que tenga registrada en la base de datos. Está claro que para poder redireccionar correctamente, el gestor tendrá que actualizar la dirección URL asociada al PURL en la base de datos. Si algunas de las direcciones URL cambian, la base de datos actualiza automáticamente todos los enlaces que se encuentren dentro de la PURL. En la página del PURL se hace referencia explícita a la persistencia de los artículos. Toda biblioteca puede pedir el software y crear su propio servidor PURL.

Tres de sus puntos débiles son:

- Las direcciones tienen que ser actualizadas;
- La dirección PURL la elige el usuario. Puesto que no pueden existir dos PURL iguales, si el sistema se difundiera como estándar, los rechazos a las peticiones serían muy largos;
- La información no se suele almacenar en el servidor PURL. Así que si una página desaparece, el PURL es inútil y la búsqueda queda vacía. Si la página en la que se encuentran los textos o artículos que queremos preservar, supone algún gasto de “alquiler” o de pago por la dirección, el propietario la borrará, perdiéndose la información.

URI - URN

Los URI (Uniform Resource Identifiers) son la tecnología standard para dirigir dentro de la Red de Redes. Incluyen las URL (Uniform Resource Locator) y los URN (Uniform Resource Name). Un URL típico se estructura así:

`http://abc/investigación/aspectos/caso_de_estudio.pdf`

`http://`, es el protocolo; `abc`, es el dominio que consiste en una traducción del número IP del ordenador; `aspectos` es el directorio en el que el `caso_de_estudio` se encuentra.

Un URN en cambio tiene más el aspecto de un isbn: 1-56592-512-2.

El problema de las URL es que indican un lugar, no un recurso. El URN es un "globally unique, persistent identifier" que se utiliza para el reconocimiento de

o el acceso a un recurso o a una unidad de información desarrollado por el Internet Engineering Task Force¹² (IETF).

Para el funcionamiento de las URN es necesario un Resolution System: un sistema que traslade las URN hacia las direcciones URL que le son específicas. Se establece un URN para un objeto digital, entonces es necesario registrarlo y guardarlo en una base de datos accesible a través de la red que puede dirigir las búsquedas hacia su localización. Una de sus características es que se pueden aplicar a cualquier objeto.

Bibliographic Protocol (BibP) and Universal Serial Item Name (USIN)

El Professor Robert D. Cameron de la Simon Fraser University, School of Computing Science, en Canada, ha trabajado mucho la cuestión de los números seriados para los documentos electrónicos. Es el creador del Bibliographic Protocol (BibP) y el Universal Serial Item Name (USIN) project at USIN.org.

BibP se define como “a web-based protocol for linking to published papers (articles, books, reports) using Universal Serial Item Names as bibliographic identifiers” (...) “to achieve persistent reference linking”.

Robert D. Cameron (Cameron 1998) apoya el desarrollo del Universal Serial Item Name (USIN) para que un sistema cree números únicos y fáciles de recordar. Se trataría de un sistema parecido a los números de identificación ya existentes y estándares (SICI, ISSN), pero mejorados:

“The USIN is intended as a vehicle for interoperability between various bibliographic citation applications, including finding citations (literature research), retrieving citations (from online sources, libraries or document delivery services), citation indexing, and citation formatting (bibliography preparation). The USIN is also intended as one possible mechanism for migrating the World Wide Web away from dependence on Uniform Resource Locators (URLs) to a system meeting the requirements for Uniform Resource Names (URNs) (...) In particular, the USIN should use mnemonic coding and be reproducible by ordinarily literate people (authors, students, librarians, law clerks, and so on) without the need for specialized coding knowledge and check-sum algorithms. The USIN system is also intended for serialized material that is not or cannot be registered with an International Standard Serial Number (ISSN)”.

Cameron hace hincapié en algo a lo que los otros investigadores no dan importancia: primar las necesidades de la gente que va a usar los USIN sobre la de los ordenadores que los procesan. Puesto que los estudiantes tendrán a menudo que apuntar USIN o teclearlos en los ordenadores, se quieren evitar los errores y facilitar la posibilidad de incluir sistemas mnemónicos para identificar a publicaciones seriadas. Sin embargo, el uso de programas como el Procite y el Endnote hace menos importante esta característica que la falta de ambigüedad a la hora de identificar un artículo: “every article must be denotable and every USIN denoting an article must denote no other article”. Si el sistema de

¹² El grupo oficial de programadores para el desarrollo de Internet.

creación de los códigos tiene la restricción de querer ser fácil o mnemotécnico, será más complicado encontrar un código único para las millones de publicaciones y artículos que van a solicitar obtener uno.

Digital Object Identifier (DOI)

La [International Digital Object Identifier Foundation](#) (IDF) y la CrossRef¹³ son dos instituciones que se basan en la tecnología del Handle System. Ambas llaman a los handles como Digital Object Identifiers (DOI). El DOI es un sistema en origen creado para identificar e intercambiar información sobre la propiedad intelectual en Internet. Se quería defender el propietario de los derechos y facilitar la automatización del comercio por la red.

El DOI System [acerca del DOI, leer (Paskin 2002)] ofrece un mecanismo único para todo tipo de información (revistas, imágenes, libros electrónicos, etc.) ofreciendo gran cantidad de información (*metadata*). Está formado por un prefijo y un sufijo. La sintaxis del DOI es: <Código del Directorio>. <Código del Propietario> /<Cadena de Sufijo DOI>, por ejemplo: 10.1000/140 Otro ejemplo: 10.1065/abc123defg•10

El [DOI System Specification](#) ha sido definido (Attanasio 2003) como “a standard identifier¹⁴ for intellectual property entities”. Se trata de una iniciativa de la Association of American Publishers (AAP) promovida para la gestión y evolución del comercio electrónico y para la gestión de los derechos de autor de objetos publicados en Internet: “DOI schemes identify documents as items owned by publishers, with numbers possibly assignable in advance of publication and independent of publication numbering¹⁵”. Los DOI y el DOI System están en funcionamiento desde Junio 1998. Ser miembro de IDF y utilizar DOI elimina la necesidad de tener tu propio handle service local, y servicios de protección de derechos que han sido estudiados para los DOI.

Las editoriales de libros electrónicos están implantando el uso de los DOI en previsión del boom del e-learning y poder así tener organizado un sistema efectivo de cobros-pagos y defensa de sus derechos.

¹³ CrossRef es un consorcio para editoriales de revistas científicas, que ha comenzado a funcionar en October 2000, es una “collaborative, reference-linking service used by journal publishers” que recoge y almacena los *metadata* de los artículos, incluidos los DOI. Puede crear DOI, y luego permite a sus miembros buscar en los *metadata* y encontrarlos.

¹⁴ “El DOI tendría así para el comercio en Internet, la misma función que tienen los códigos de barras para el comercio de productos tangibles: facilita la interoperabilidad entre sistemas de información”..”no sólo (para todo tipo de objetos digitales, sino también para toda entidad IP, incluyendo los resúmenes, libros, etc” (Traducción del autor). Attanasio, P. (2003). [Standard identifiers for learning objects: the DOI case](#). E3 - Education, Europe And E-Government, Roma, www.medra.org.

¹⁵ “The reliability and accuracy of the DOI give it the ability to become a form of electronic ISBN/ISSN but with the added value that the metadata it delivers can focus on publisher issues primarily, rights management and electronic commerce”.

Al DOI se le puede considerar como una implementación realizada de los identificadores permanentes (i.e. Handles) por parte de una comunidad de intereses, las editoriales.

La idea es que el DOI coexista con los identificadores existentes hasta ahora y utilizados para la prensa escrita (i.e. ISBN, ISSN, SICI). En determinados casos puede incluir estos identificadores en la secuencia del DOI. El sistema se está diseñando para que estén solucionadas ciertas cuestiones importantes como la variación en tipos de contenido; diferentes manifestaciones de un mismo objeto; cambio en la propiedad del objeto.

La AAP ha aceptado la regla general de que un DOI debería reflejar una sola URL, y que el poseedor de los derechos de un objeto es el único que puede “crear” una DOI (“is the only one authorized to issue a DOI”). Se plantea el problema de los objetos con varios propietarios para diferentes derechos. Parece que la International DOI Foundation limitará el derecho de crear DOI a las editoriales y a organizaciones similares.

En la Public Library of Science, PLoS Biology Author Guidelines se escribe para los artículos a publicar: “Use of a DOI number to the full-text article is acceptable as an alternative to or in addition to traditional volume and page numbers.”

Hemos señalado que el DOI ha sido creado por las editoriales para defender sus intereses y es sólo “una herramienta para describir los objetos identificados” (Attanasio 2003), aun no soluciona realmente nuestro problema puesto que su objetivo no es el de asegurar que el contenido se mantenga, sino poder encontrar la mejor información disponible. Respecto a esto, parece que las posibilidades de los DOI son mayores que las de un entorno URN y el Sistema Handle: a parte proporcionar mecanismos para crear y resolver los únicos identificadores, el sistema DOI se puede extender para incluir la base de datos que contiene el material, o la información relacionada, que deriva de una búsqueda DOI.

Al hablar de las funcionalidades del DOI (Attanasio 2003b), se comenta el “voluntary deposit”.

[mEDRA multilingual European DOI Registration Agency](#)

Mencionar la reciente creación de mEdra como la plataforma europea para la creación y el crecimiento de los DOI. Se trata de un proyecto cofinanciado por la Comisión Europea a través del E-content Program. Será una empresa comercial desde julio 2004.

El problema de estas herramientas listadas es que garantizan la persistencia del nombre del objeto de la última dirección, sin embargo no garantizan la permanencia del recurso.

Veamos alguna propuesta que intenta solucionar el problema de la permanencia de los recursos.

Propuestas para la permanencia de documentos para referencias

Hay dos propuestas que queremos destacar. La primera es la de Mark Bide. En su artículo titulado *In Search of the Unicorn* (Bide 1998), Bide prevé un futuro en el que habrá registros centrales digitales. Cada trabajo tendrá un identificador único, que las bibliotecas podrán obtener y se podrá dar acceso a materiales previamente digitalizados. Bide propugna por, en aquel entonces, recién creado DOI. El de Bide se trata de un modelo muy vigente. En un informe (M Bide 1999), Bide y otros investigadores defienden la necesidad de un registro voluntario previo al obligatorio/legal para probar el funcionamiento. En las conclusiones del informe se dice:

“To this end, we recommend the establishment of a joint committee, made up of technical representatives of the deposit libraries and a broad cross section of the electronic publishing community to attempt to find mutually acceptable solutions to the questions posed. This committee should be established under the auspices of Book Industry Communication which is uniquely placed to ensure that the interests of both communities are properly served”. (...) “The Kenny report¹⁶ concludes that legal deposit represents the best method of ensuring the development of a comprehensive national digital archive. However, voluntary or legal deposit is a very different instantiation of “digital preservation” than that exemplified in other “digital preservation” schemes. In particular, the relationship between the creators of the material (publishers) and its preservers (the deposit libraries) is very different in the national deposit context as compared with most other digital archives, where preservation issues *inform* decisions taken about the way data is created”.

Es necesario que se cree un organismo supranacional y varios nacionales, para crear una base de datos que mantenga todos los identificadores con sus datos asociados completos. Para hacer una comparación, se trata de una especie de biblioteca nacional virtual.

Robert D. Cameron (Cameron 1997) ha propuesto “a universal, Internet-based, bibliographic and citation database would link every scholarly work ever written - no matter how published - to every work that it cites and every work that cites it”. Cameron propone una base de datos creada con los mismos criterios en las diferentes instituciones que producen trabajos científicos (Universidades, Institutos de Investigación,...) que permitiera la búsqueda cruzada de todos los datos. Señala como clave para el éxito de este modelo, que sean los gobiernos los que obliguen a través de alguna norma que se aporte la información, para asegurar así la universalidad de los datos: “Finally, development of data by the originating institutions¹⁷ can be carried out even if some publishers refuse to participate; in this way, the originating institutions regain a measure of control over their publications to balance the loss of

¹⁶ Report of the Working Party on Legal Deposit. This is available at <http://www.culture.gov.uk/LDWGRPT.HTM>

¹⁷ Las universidades, por ejemplo.

copyright typically required for journal publication”. Las Instituciones deberían coordinarse para que la información no estuviera sólo parcialmente disponible. Otra propuesta de Cameron es el modelo de “a semi-universal citation database. Esta base de datos sería igual que la anterior, pero sólo incluiría todos las citas posteriores a una determinada fecha de comienzo. Los trabajos más antiguos estarían presentes en la base de datos con su información bibliográfica, y sólo si han sido citados por publicaciones posteriores a la fecha de inicio. Se trata, en conclusión, de un proyecto menos ambicioso que el precedente. Por otro lado, parece que la preocupación de Cameron se centra particularmente en la posibilidad de encontrar los enlaces y las citas, más que la permanencia del contenido, algo que parece dar por descontado.

En otro lugar hemos comentado que el problema que nos estamos planteando no incluye el de la cantidad de información falsa, poco precisa o sin valor que existe en la red, sin embargo hay aspectos que nos afectan. Cameron sugiere que se aplique un filtro:

“One could also use citation counts for filtering; papers would only be selected if they had been cited a certain number of times (or by a certain number of distinct authors). This could be particularly useful for general or initial reading in an area; effort could be concentrated on those papers deemed important by the research community through their cumulative citation counts”.

De momento, para poder implementar este sistema, sería necesario que un programa pudiera leer el número DOI de la bibliografía, y esa información se tratara. La solución que algunas empresas están implementando puede ser más simple. Actualmente, hay páginas web con artículos gratuitos que ofrecen facilidades para encontrar aquellos artículos que han sido “bajados” un mayor número de veces, considerándolos así como a cuanto más solicitados, más interesantes.

Un proyecto a seguir es el que se está llevando a cabo en Alemania. Se van a aplicar los DOI para identificar de forma duradera los datos presentes en la German National Library of Science and Technology (TIB), la biblioteca de ciencia y tecnología más grande del mundo:

“This use of DOI will provide for the effective publication of primary data using a persistent identifier for long-term data referencing, allowing scientists to cite and re-use valuable primary data. The DOI's persistent and globally resolvable identifier, associated to both a stable link to the data and also a standardised description of the identified data, offers the necessary functionality and also ready interoperability with other material such as scientific articles.”¹⁸.

Otra iniciativa de gran interés y que ya se ha asentado en el campo de las ciencias de la naturaleza, es el PubMed Central. Se trata de un archivo digital para las revistas de ciencias naturales, desarrollado y gestionado por el National

¹⁸ Ver: "[Project announced to develop DOIs for scientific data: German National Library of Science and Technology joins IDF](#)

Center for Biotechnology Information (NCBI) en la U.S. National Library of Medicine (NLM)”. El acceso a PubMed Central es libre y sin restricciones. La función de PubMed Central es el de ser una base datos para localizar “relevant articles and, in many cases, link directly to a publisher's site for the full text”. Aunque la participación de editoriales es voluntaria, se restringe a las revistas que cumplan con ciertos estándares.

En consonancia con el problema planteado, se dice:

Many journals already have online publishing operations and there is a growing tendency to publish material online only, to the exclusion of print. This literature must be preserved in a form that ensures open access to it over the longer term. This is what NLM has undertaken to do. We invite all journals to join those that have already committed to creating this resource for people all over the world.¹⁹

Otra iniciativa interesante y relacionada con nuestro problema es la del SEP (Deposit System for Electronic Publications) (Werf 2000) del NEDLIB.

Modelo para la permanencia de documentos en Internet

Una vez que hemos hecho un repaso por el estado de la situación, tenemos datos suficientes para poder ofrecer una propuesta a nuestro problema.

Los sistemas vistos avanzan en la dirección correcta para que se asegure la permanencia de enlaces y contenidos. En la misma línea de las propuestas de Cameron y Bide, parece necesaria la creación de una base de datos que mantenga y proteja los documentos²⁰.

Las entidades de gestión, puesto que se trata de Internet, deberían ser internacionales, teniendo como criterio de división el idioma. Así, en el ámbito hispano, la Biblioteca Nacional de España podría crear una filial (llamémosle *Depósito de Registros Virtuales – DRV*²¹) con acuerdos con las otras bibliotecas nacionales de países hispanohablantes.

Por otro lado, parece sensato proponer que sean los Estados quien regule el funcionamiento de esta base de datos de documentos virtuales puesto que se trata de información de interés general.

El sistema voluntario sería difícil de implementar, puesto que entes como las editoriales tendrían un aumento en los gastos (i.e. personal para tratar e introducir los datos) a los que estarían difícilmente dispuestos sin una

¹⁹ Algunas bases de datos que pueden ser un ejemplo del posible desarrollo futuro son: [FlyBase](#); [InterPro](#); [LocusLink](#); [OMIM](#)

²⁰ Hay muchos temas que surgen aquí para la discusión: de un objeto que evolucione en el tiempo por su alto grado de interactividad, ¿qué hace falta registrar?; de los objetos que tienen imágenes en vídeo, audio, etc. , ¿qué es necesario registrar? Ver M Bide, E. J. P., A Watkinson (1999). Digital Preservation. An introduction to the standards issues surrounding the deposit of non-print publications, The Library and Information Commission: 45.

²¹ Ver el interesante artículo de Werf, T. v. d. (2000). The Deposit System for Electronic Publications: A Process Model. [NEDLIB Report Series](#). Amsterdam, NedLib Consortium: 62. sobre cómo implementar un sistema de biblioteca virtual.

contrapartida monetaria, aunque Bide señale ciertos intereses por parte de los editores en realizar esta labor (M Bide 1999):

“Publishers also recognize, as we have mentioned above, that deposit might satisfy one of their direct interests, namely the ability to meet authors’ concerns about a permanent archive of their work. More generally, as contributors to the information chain, they are not entirely uninterested in, or unsympathetic to, the development of a national archive which, after all, can prove as fertile a resource for the development of publishing projects as for scholarly research”.

Parece que, como sugería Bide, se podría proponer un sistema completamente voluntario en la primera fase de pruebas. Después, ciertos materiales (estamos hablando de trabajos de transferencia de información) se deberían registrar obligatoriamente; posiblemente, los que fueran a venderse de una forma directa o indirecta (a través de suscripción a una página o a un servicio). El resto podría seguir registrándose de una forma voluntaria.

El DRV debería ser un ente público o semipúblico. La gestión podría proporcionar beneficios si el registro de documentos fuera bajo pago de una pequeña cantidad. Esta cantidad serviría para el mantenimiento y expansión de la base de datos; además de filtrar los trabajos sin pretensión de calidad o de permanencia.

Veamos el proceso: todo comienza cuando una persona que tiene algo que decir de posible interés, lo escribe. No es necesario que se trate de un investigador.

En este momento pueden darse los siguientes escenarios: 1. Registra en la base de datos. A la hora de registrarlo tendrá que enviar una copia del documento al que la Base de Datos asignará automáticamente un NI (Número Identificativo; podría ser un DOI). El documento se almacena con el número correspondiente²². Posteriormente el documento se puede publicar en Internet con el NI. 2. Puede mandar el documento a una revista que se encargará de registrarlo y publicarlo; 3. Puede publicarlo en Internet sin registrarlo; 4. Puede mandarlo a una institución que esté autorizada para crear NI. Se nombrarían unas entidades de reconocido prestigio y que garanticen el mantenimiento y la permanencia de los fondos, con poder para asignar números de identificación: las Universidades, por ejemplo. Esta institución se encargará de archivar el documento y emitir el número. Luego deberá transmitir al órgano central todos los documentos con el correspondiente NI creado. Es importante que el formato en el que se trabaje sea el mismo.

Surgen aquí algunas cuestiones de importancia: 1. La base de datos pública (el DRV) no puede soportar el peso de todas las búsquedas y solicitud de documentos. Es necesario diseñar un sistema por el que el DRV sólo envíe aquellos documentos que están perdidos y, si no lo están, ofrezca la posibilidad de encontrar el documento en un servidor “mirror” (el de la Universidad que le ha enviado un determinado documento, por ejemplo). Al mandar una copia a la base de datos, se le asigna automáticamente un número, comparable a un

²² El almacenaje del material digital tiene muchas dificultades que han sido ya ampliamente estudiadas.

número de DNI. El punto es: si hay más de un servidor en el que se encuentra el documento, ¿Hacia dónde dirige ese ID? Teniendo en cuenta que va a existir un archivo central, se debería descentralizar la responsabilidad de mantener el sistema ágil. Por ello los enlaces llevarán a los servidores de una entidad autorizada que ha creado y almacenado el ID. Podría haber posiblemente la posibilidad de elegir entre diferentes *mirrors*. 2. Se tiene que plantear el sistema de cobro antes mencionado para que se trate de un modelo equilibrado. Este tema tiene particular relevancia cuando estamos hablando de derechos de autor y de intereses de editoriales. Los DOI ofrecen, por ejemplo, la posibilidad de gestión de derechos de autor. Se puede utilizar un sistema de *stamping technologies* que graba la hora de registro de cada documento. De hecho, este número identificativo certificaría la autoría de un determinado documento en Internet. No voy a entrar en el tema de los derechos de autor que tanto han sido tratados y para el cual me falta información legal y tecnológica²³. 3. El número generado sólo vale para las publicaciones *on line*.

El número iría en las citas y a través de él se localizaría el documento. Después de un período de carencia para que las editoriales exploten los artículos de una forma privativa, los documentos registrados serían públicos. En PubMed Central utilizan un sistema similar, sin embargo las editoriales mantienen los derechos y esto puede hacer que los artículos sean retirados cuando a aquellas les pueda parecer oportuno: “Although immediate open access to all content is most desirable, a journal may delay release of its full text in PMC for some period of time after publication” (...) “Copyright remains with the journal publisher or with individual authors, whichever is applicable”. Los artículos deben tener un régimen de derechos de autor diferente al de los libros.

Las editoriales de revistas a lo mejor verían disminuidos sus ingresos, pero sobrevivirían puesto que el verdadero valor añadido que ofrecen es el seleccionar la información mejor y más actual, y presentarla de una manera ordenada.

Todo el sistema se limitaría a los artículos publicados en Internet. Si se publican en formato impreso, el proceso sería el seguido hasta ahora. No estoy de acuerdo con los autores que sostienen que los artículos se deberían registrar sólo una vez, físicamente o digitalmente: “The voluntary code of practice states that ‘publications which appear with substantially identical content in more than one medium only need to be deposited in one medium’ “(M Bide 1999) De hacerse así, aunque parezca paradójico, el sistema se complicaría en gran medida. Lo que se publica en Internet debería depositarse y registrarse en Internet, aunque sí valdría el registro del producto físico para certificar la autoría. Paskin (Paskin 2002) recoge este mismo problema al decir que

“While it may often make sense for different manifestations of the same intellectual content to have different identifiers, the underlying abstract work, the journal itself, is the same. This makes decisions on what is and

²³ Por ejemplo, ver Chilvers, A. (2002). "A possible way forward with specific reference to the UK." *Journal of Documentation* **58**(2): 146-174.

cooperación a nivel mundial entre las bibliotecas”, (...) “Promover las iniciativas que faciliten el acceso, incluido el acceso libre y asequible, a los periódicos y libros de libre acceso, y a los archivos de libre acceso que contienen información científica” (...) “Definir mecanismos seguros para el almacenamiento y el archivo de los documentos y otros registros electrónicos de información”. (2003b)

Nuestro país debería estar en la vanguardia en estas propuestas. Por ahora está en la cola.

Bibliografía

(2002). Media-related Identification and Metadata Standards, EDRA – European DOI Registration Agency.

(2003). Declaración de Principios: Construir la sociedad de la información: un desafío mundial para el nuevo milenio. UIT (Naciones Unidas). Ginebra: 11.

(2003b). Plan de Acción. UIT (Naciones Unidas). Ginebra, WSIS: 18.

Attanasio, P. (2003). Standard identifiers for learning objects: the DOI case. E3 - Education, Europe And E-Government, Roma, www.medra.org.

Attanasio, P. (2003b). DOI: the next phase. The mEDRA applications. 25th International Supply Chain Specialists Meeting, Frankfurt, www.medra.org.

Bide, M. (1998). In Search of the Unicorn. The Digital Object Identifier from a User Perspective. BNBRF Report. London, Book Industry Communication: 40.

Cameron, R. D. (1997). A Universal Citation Database as a Catalyst for Reform in Scholarly Communication.

Cameron, R. D. (1998). "Towards Universal Serial Item Names Journal of Digital Information." School of Computing Science, Simon Fraser University 1(3).

Chilvers, A. (2002). "A possible way forward with specific reference to the UK." Journal of Documentation 58(2): 146-174.

Dyson, E. (2003). "Online Registries: The DNS and Beyond..." Release 1.0 21(8).

Eco, U. (1988). Opera aperta. Milano, Gruppo editorial Fabbri, Bompiani.

Gilheany, S. (1998). Preserving Information Forever and a Call for Emulators. Digital Libraries Asia 98: The Digital Era: Implications, Challenges & Issues, Singapore, www.ArchiveBuilders.com.

M Bide, E. J. P., A Watkinson (1999). Digital Preservation. An introduction to the standards issues surrounding the deposit of non-print publications, The Library and Information Commission: 45.

Martínez, D. (2003). "La crisis de las revistas científicas y las nuevas oportunidades de Internet." Revista Telos(56. Segunda Época).

Paskin, N. (2002). Digital Object Identifiers. ICSTI Seminar: Digital Preservation of the Record of Science, IOS Press.

Payette, S. (1998). "Persistent Identifiers on the Digital Terrain." RLG DigiNews 2(2).

Polanco Masa, A. (2002). "Historia e Internet. El incierto legado de la Cultura Digital." Panacea, Revista de Historia de la Ciencia(3).

Salomon, D. J. (1999). "Is it Time to take the paper out of Serial Publication?" Medical Education On Line 4(7).

Sierra Bravo, R. (1999). Tesis doctorales y trabajos de Investigación Científica. Madrid, Editorial Paraninfo.

Thomas, D. (2003). Digital Preservation at the National Archives.

Werf, T. v. d. (2000). The Deposit System for Electronic Publications: A Process Model. NEDLIB Report Series. Amsterdam, NedLib Consortium: 62.