
Preservación digital

Digital preservation

Carlos G. FIGUEROLA (1), Raquel GÓMEZ DÍAZ (2),
José Luis ALONSO BERROCAL (3) y Ángel ZAZO RODRÍGUEZ (4)

(1) Facultad de Traducción y Documentación, Francisco Vitoria 6-16 37008 Salamanca, figue@usal.es
(2) rgomez@usal.es (3) berrocal@usal.es (4) zazo@usal.es

Resumen

Partiendo del gran volumen de la información que se encuentra en formato digital y de la heterogeneidad en cuanto al modo de producción, soportes, y dispositivos que facilitan su acceso, en el artículo se señalan algunos de los problemas que plantea la información digital así como iniciativas que tratan de dar respuesta al problema de su preservación.

Palabras clave: Preservación digital. Problemas. Soluciones

Abstract

Based on the large volume of digital information and the heterogeneity in the production, media, and devices that facilitate access, the article highlights some of the problems of digital information as well as the main initiatives that seek to address the problem of preservation.

Keywords: Digital preservation. Problems. Solutions.

1. Introducción

Cada vez es mayor el volumen de información disponible en formato digital. Una parte creciente es sólo digital: se produce con medios digitales y no pasa nunca a un soporte físico. Paralelamente, asistimos también a un proceso inverso: información disponible en soporte físico convencional (por ejemplo, papel) pasa a convertirse en digital, ya sea como estrategia de conservación, o simplemente para facilitar su uso. Esta heterogeneidad no es sólo por su origen, también afecta al modo en que la información se codifica, se estructura o se plasma en un documento y, por tanto, también a cómo se codifica y se usa.

Partimos de que el mismo concepto de 'digital' es impreciso: en algunas de sus formas requiere un soporte físico (por ejemplo un CD/DVD); en otros casos nos encontramos con documentos que sólo existen en Internet, y dentro de éstos, muchos ni siquiera existen como tales, sino que se componen 'al vuelo' y a demanda a partir de diversos elementos almacenados bajo diversas formas (por ejemplo, las páginas web dinámicas).

El predominio creciente de lo digital ofrece una serie de perspectivas imposibles e impensables antes, pero plantea también numerosos interrogantes, así como la necesidad de resolver determinados problemas. Por ejemplo en el terreno del acceso a la información que contienen los documentos digitales. Al contrario de lo que

ocurre con un documento convencional (texto escrito sobre papel), el acceso al contenido de uno digital requiere intermediarios: una (o más) máquinas y programas o software que permiten acceder a su contenido. Tanto unos como otros evolucionan de manera muy rápida y se vuelven obsoletos en espacios muy cortos de tiempo. Las máquinas y los programas que utilizamos hoy, probablemente no se usarán dentro de unos pocos años. La pregunta es si vamos a seguir pudiendo acceder a información que se plasmó para ser usada con esas máquinas y esos programas.

Algo parecido sucede en el terreno de la autenticidad y cuestiones anejas. La facilidad de replicación permite también el surgimiento de apócrifos en todos los sentidos. Es cierto que los sistemas criptográficos actuales (firma digital, registro digital, etc.) garantizan, cuando se aplican, autoría, fecha, integridad y demás, pero no es menos cierto que la próxima generación de ordenadores sobrepasará las posibilidades de la criptografía actual, con lo que tales garantías dejarán de serlo.

Este artículo pretende señalar algunos de estos problemas y afrontar una reflexión sobre ellos, más en la idea de llamar la atención sobre esas cuestiones que en ofrecer soluciones concretas, ya que éstas están, hoy por hoy, creemos, lejos de nuestro alcance. Así, este trabajo se estructura en dos grandes bloques. En el primero de ellos se describen algunos problemas significativos surgidos en torno a lo que podemos llamar,

de forma genérica, información digital; muchos de éstos vienen dados por el crecimiento vertiginoso de este tipo de información. Pero no es sólo una cuestión cuantitativa; así, se abordan en esta sección cuestiones como los cambios en el proceso de producción documental, así como en el de la difusión de la información digital; la desaparición de los soportes físicos y la proliferación de documentos llamados dinámicos. Se abordan también problemas relacionados con el acceso a la información, tanto desde el punto de vista de los intermediarios de todo tipo de procesos como desde el de la mera localización de la misma; y también aspectos como los derivados de cuestiones como la copia, autenticación, derechos de autor, etc. en el terreno digital. En el segundo bloque del artículo se repasan algunas de las soluciones emprendidas o aplicadas desde diversos ámbitos; todas ellas son parciales, en el sentido de que abordan sólo aspectos determinados o concretos. De la mayor parte se desconoce aún su recorrido real, pues su tiempo de vida es extremadamente corto; pero el análisis de ellas puede ayudar a comprender las dimensiones reales de los problemas que aquí tratamos.

2. Algunos problemas en torno a la información digital

2.1. El desarrollo de la información digital

Aunque resulta difícil de cuantificar el volumen de información que hay en Internet, según el informe *El Diverso y Explosivo Universo Digital*: Un pronóstico actualizado del crecimiento de la información mundial hasta el 2011 (Grantz, 2008) realizado EMC e IDC, en la actualidad hay unos 487.000 millones de gigabytes, el triple del volumen que había en 2007, y según las previsiones de EMC Consulting en poco más de año y medio este volumen se duplicará. Se estima que el crecimiento anual de la información en 2007 fue aproximadamente de un 60% llegándose a los 281,000 millones de gigabytes, y se espera que en 2011 se llegue a unos 1,800 exabytes. Uno de los datos más relevantes de este estudio es que la cantidad de información que se genera sobre una persona es mayor a la cantidad de información creada por la misma, lo que se conoce como *sombra digital*.

De otro lado, no es sólo un cuestión de cantidad; el propio sistema de producción documental ha experimentado una transformación sin precedentes. Naturalmente, lo más inmediato en este sentido es acudir a los aspectos meramente tecnológicos: la imprenta tiene poco que hacer frente a un procesador de textos, igual que la escritura manual tampoco fue rival para

la primera. El proceso de producción es ahora mucho más rápido, pero también es mucho más barato y está al alcance de mucha más gente. Tenemos, pues, a muchos más productores de documentos, que además generan documentos a mucha más velocidad. Por otro lado, el modo de producción genera fenómenos curiosos, pero también significativos: es posible la producción automática de documentos, así como, más modesta pero, por ahora, más frecuente, la semi-automática: lo que todos conocemos como *cut & paste*. Este fenómeno permite, por ejemplo, la replicación de información en documentos que formalmente son diferentes, incluso muy diferentes.

Los cambios en el modo de producir los documentos afectan a cosas como la autoría de los mismos: tenemos documentos colaborativos, a veces por varios autores concretos, pero otras muchas por innumerables colaboradores a menudo anónimos. Quizá el ejemplo más claro es el de la Wikipedia (Lih, 2004), donde gracias al trabajo de muchas personas se puede encontrar información prácticamente de cualquier tema. Este ejemplo no es el único, sino que cada vez hay más plataformas colaborativas que tienen como fin la difusión libre del conocimiento mediante el trabajo arduo de millones de usuarios que a diario lo están actualizando.

En este sentido también es reseñable el nuevo modo de crear literatura, así, algo que nació de la mano de los juegos de rol donde la historia la escribían los personajes que vivían cada aventura, ahora Internet también ofrece la posibilidad de participar en la creación de distintas obras. En este sentido encontramos distintas iniciativas en la red (1).

Asociado a la producción también están las versiones, revisiones o ediciones de un producto. Es decir, el estado en el que se encuentra en un momento dado en su desarrollo o modificación. Esto se denomina control de versiones, y permite conocer la gestión de los diversos cambios que se realizan sobre los elementos de algún producto o una configuración del mismo. Los sistemas de control de versiones facilitan la administración de las distintas versiones de cada producto desarrollado, así como las posibles especializaciones realizadas por ejemplo, para algún cliente específico.

Uno de los elementos que nos permite ver la evolución de la información es el control de versiones. Esta práctica que es muy habitual en informática, donde a través del código fuente podemos ver los cambios que han tenido los distintos programas (Collins-Sussman, 2008), no lo es tanto en otros ámbitos donde llega un

momento en el que no podemos adivinar fácilmente quién, ni cuándo ha modificado un determinado documento. Es cierto que puede resultar poco útil saber en un trabajo colaborativo qué ha aportado cada persona, pero hay en entornos donde resulta muy importante saber si alguien ha modificado un documento. Así, en el ámbito archivístico, esto adquiere especial relevancia, ya que en todo momento debemos controlar la manipulación de los documentos. En los documentos de archivo se hace a través de la trazabilidad del documento (también llamada pista de auditoría) donde, además de indicar por dónde ha pasado el documento, se deben señalar los cambios que ha sufrido. Así, para la *UNE-ISO 15489* este proceso aporta seguridad, ya que garantiza que el documento ha sido manipulado exclusivamente por quien puede hacerlo. De este modo, junto con el documento es necesario guardar también la información generada a partir de la vida de ese documento. El problema se presenta cuando no somos capaces de documentar esos cambios.

Fuera del mundo de los programas informáticos y de los archivos resulta complejo encontrar iniciativas que nos permitan ver estas evoluciones de los documentos, páginas web...; y, aunque los editores de texto suelen incluir una aplicación que permite mostrar las diferencias entre varios documentos, esta herramienta es limitada y sólo permite la comparación señalando las diferencias.

2.2. El acceso al documento digital

Los problemas relacionados con el acceso de las personas a la información digital, en cualquiera de sus acepciones, son variados, y probablemente sólo somos conscientes ahora de una lista reducida de los mismos.

Uno de los más señalados es el de la cantidad de elementos requeridos para acceder a esa información (Boorgman, 2000). En el caso de los documentos tradicionales, básicamente texto escrito sobre papel, por poner el caso más frecuente, todo lo que se necesita es el uso de nuestros propios sentidos (la vista) y el conocimiento de unas técnicas que consideramos básicas: la lectura. Ciertamente es que, en el caso menos frecuente de querer acceder a documentos convencionales de determinadas épocas históricas es preciso manejar conocimientos adicionales especializados (paleografía, etc.); pero no es menos cierto que ésta es una situación poco frecuente que se produce en ámbitos muy especializados. Dicho de otra manera, el acceso a la información del documento convencional no exige más intermediario que el saber

leer; y éste es un tipo de destreza generalizado que forma parte del bagaje cultural normal, incluso más allá de lo que llamamos cultura occidental.

El problema es qué ocurre con aquellos países cuyo índice de desarrollo humano según el Informe del PNUD está en los puestos inferiores. Si cualquiera de estos países tiene pocos recursos, éstos no se invertirán en infraestructuras tecnológicas, con lo que la distancia en desarrollo entre los países que ocupan los primeros puestos del ranking y los últimos cada vez es mayor (2).

Además debemos tener en cuenta que el acceso a la información digital requiere más cosas, no sólo conocimientos, sino también otro tipo de intermediarios materiales: precisa de máquinas o dispositivos, en el caso más típico, ordenadores (aunque podrían ser otros dispositivos), y también programas o aplicaciones en un sentido amplio. Así es que, de un lado, necesitamos destinar más recursos de los que no siempre es fácil disponer, además, el uso de tales elementos muchas veces requiere un nivel de destrezas superior al de la mera lectura. Además, tanto los elementos físicos (hardware) como los lógicos (software) que se necesitan son complementarios entre sí. Su evolución es autónoma y no siempre en la misma dirección.

Además, evolucionan de forma muy rápida; mucho más rápida de lo que ha evolucionado la escritura convencional (Mens, 2005). De manera que las máquinas que hoy usamos para acceder a la información digital no tienen nada que ver con las que había hace 10 ó 15 años; ni probablemente se parezcan a las que habrá dentro de otros 10 años.

En el campo de los elementos lógicos o software, el panorama es, incluso, un poco más complicado. En efecto, un programa no suele ser un elemento aislado, sino que depende, como poco, de un sistema operativo, que a su vez suele constar de varias capas; pero generalmente también hará uso de otros componentes software producidos por terceros (librerías, pluggins, etc.). El resultado es que todos estos componentes son necesarios para que ese programa pueda funcionar y permitimos acceder a la información digital producida con él.

Esto plantea el interrogante de si dispondremos dentro de unos pocos años de los intermediarios tecnológicos necesarios para acceder a la información digital que estamos generando ahora mismo. De hecho, ya tenemos problemas para acceder a la información digital producida no hace tanto tiempo; es difícil encontrar, por ejemplo, lectores de disquetes de 5.25 pulga-

das, y los más conocidos de 3.5 pulgadas van camino de desaparecer. Mucho más difícil es encontrar dispositivos capaces de acceder a las numerosas variedades de discos ópticos que había hasta que se generalizó el CD (del cual hay, por cierto, diversos formatos, algunos no tan fáciles de leer); y éste, el propio CD, probablemente tiene los días contados. A esto tenemos que añadir la pugna de las empresas por imponer sus dispositivos (por ejemplo, la lucha entre DVD y BlueRay)

En el mismo sentido, y si pensamos en los aspectos lógicos, no hay más que recordar programas que ya no se utilizan, pero que, además, aunque recuperásemos una copia de los mismos, en muchos casos no podríamos hacerla funcionar con los sistemas operativos actuales, ni tampoco con las tarjetas gráficas que usamos hoy en día. Los formatos en los que muchos de esos programas guardaban la información ya no son reconocidos por los que usamos hoy en día. La mera codificación de caracteres ha evolucionado en un par de decenas de años; las ampliaciones del código ASCII que florecían no hace tanto dejaron de usarse, por no hablar de códigos diferentes, como el EBDIC, que en su día fue muy utilizado. Incluso los diferentes códigos ISO 8859 están siendo claramente desplazados en favor de Unicode (el cual, por cierto, puede ser codificado con una gran variedad de esquemas distintos) (Marcus, 2009). Todo esto sin salirnos del ámbito de los ordenadores domésticos o de escritorio.

Por otra parte, el acceso a la información digital tiene otras aristas; entre ellas, la derivada de la gran cantidad de información disponible: ¿cómo encontrar la información realmente relevante frente a cada necesidad informativa? La abundancia de información disponible ha ido desviando el acento en la precisión frente a la exhaustividad. En muchos campos, cuando hace unos años la posibilidad de ofrecer toda la información disponible —o lo más próximo a esto— sobre un asunto determinado era un valor en sí mismo, hoy tendemos, en muchos casos, a valorar precisamente los sistemas que podan o apartan lo menos relevante, en favor de menos información, pero más precisa. Hay tanta información que, en muchas ocasiones, la exhaustividad es un obstáculo: es difícil digerir tanta información.

No es éste el lugar donde abordar la ya vieja polémica de recuperación de información frente a la recuperación de datos (Rijsbergen, 1979); polémica que, aunque vieja, parece resucitar en los últimos tiempos: a fin de cuentas, los metadatos no hacen sino aplicar un modelo propio de la recuperación de datos (Kim, 2005).

2.3. La preservación de la información en la era digital

En este panorama de abundancia, diversificación y heterogeneidad resulta complicando enfrentarse a la conservación (Kuny, 1998). Quizá la primera pregunta que debemos hacernos es ¿qué conservar? (IFLA and IPA, 2002). Pudiera parecer que ante la gran capacidad de los dispositivos y el bajo coste de los mismos podríamos tratar de conservarlo todo, pero esta solución no es viable (recordemos las tasas de crecimiento y los problemas añadidos como la duplicidad de la información). Tampoco es una solución realista, sobre todo pensando en que la conservación de la información digital implica una actitud dinámica frente a la pasividad de la conservación de otros materiales.

A lo largo de estos pocos años (las primeras experiencias de conservación son de los años 90), se han ido proponiendo alternativas distintas que pasan desde la congelación de la técnica y de los documentos creados con la misma, aislándolo de las posibles evoluciones, en museos tecnológicos; a aquellas alternativas que defienden el mantenimiento de la información (Gómez, 2009), aunque para ello haya que variar la manera de codificación de la misma, que es al fin y al cabo lo que conseguimos con la migración (Asproth, 2005). No es nuestro objetivo analizar en que consisten estos métodos.

Tampoco es cuestión de ser más conservador o más innovador, el problema es que la mayor parte de las acciones que se realizan en materia de conservación no son reversibles y que una determinada actuación o la falta de ella pueden provocar la pérdida irremediable de información más o menos valiosa.

Otra cuestión que tampoco podemos dejar de lado es lo relativo a quién tiene la responsabilidad de conservar. Este problema está muy relacionado con el mundo de los archivos. En una época donde se trata de implantar la administración electrónica, parece razonable pensar que la propia administración tiene ya resuelta esta cuestión; pero, aunque la Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos aborda el tema de la conservación, aún no ha transcurrido el tiempo necesario para verificar si los mecanismos son los adecuados.

2.4. Otros problemas: derechos de autor, libre copia y derivados

La conservación se pone en relación con otra serie de derechos como es el caso del derecho a la comunicación. De este modo mientras que

cuando conservamos un documento en soporte tradicional, lo que hacemos es transferir el soporte (contenido y continente están unidos); en el caso de los documentos digitales, al transferir la información, estamos reproduciéndola o incluso cambiando la manera de codificación de la misma, lo que podría chocar con las leyes de protección de los derechos de autor, pensadas más para otro tipo de entornos.

Este es uno de los muchos casos en los que la legislación va por detrás de la realidad y trata de trasladar patrones de otras épocas a los nuevos entornos, pero es de esperar que, al igual que los distintos países están adaptando su legislación de Depósito Legal, en el caso de los derechos de autor también se haga.

Por el momento contamos con la Directiva 2001/29/CE del Parlamento Europeo, sobre armonización de ciertos aspectos de derecho de autor y derechos vecinos en la sociedad de la información, donde se prevé la excepción a los derechos de reproducción para aquellos casos en los que se trate de reproducciones realizadas por bibliotecas accesibles al público, establecimientos de enseñanza, museos o archivos que no persigan fines comerciales o económicos directos o indirectos (Cordón, 2007).

3. Algunas iniciativas interesantes

Parece evidente que no hay soluciones mágicas ni situaciones ideales, pero no podemos obviar que la información digital es un hecho que no puede parar; de manera que, aunque no sean las mejores soluciones, en determinados casos no queda más remedio que aplicar medidas de urgencia que permitan seguir adelante, a pesar de la incertidumbre del futuro. En este sentido, surgen numerosas iniciativas que intentan abordar algunos de los aspectos reseñados más arriba. Aquí citaremos algunas de las más señaladas, apuntando también sus carencias.

3.1. Formatos abiertos

Uno de los problemas más importantes es el de los formatos de los documentos digitales. Durante mucho tiempo, el formato de un documento venía proporcionado por el programa que lo producía; si un documento era producido, por ejemplo, por un determinado procesador de textos, el formato era el propio o propietario de ese programa. Como la mayor parte de dichos programas exigían el pago de derechos o licencias de uso y se protegían, de una u otra forma, contra copias, esos formatos eran mantenidos en secreto. El resultado es que los documentos con ese formato no podían leerse con otros programas diferentes. Como dichos programas

y sus formatos eran propiedad de empresas comerciales, no resultó infrecuente que dichas empresas desapareciesen, y con ellas la posibilidad de acceder a documentos en sus formatos propietarios.

La solución evidente es utilizar formatos cuyas especificaciones sean conocidas y públicas en su totalidad; esto abre la posibilidad de que diferentes desarrolladores de software puedan escribir programas capaces de entender los documentos codificados en tales formatos. El uso de formatos abiertos no está contrapuesto necesariamente a la visión de libre empresa como motor de los avances tecnológicos; simplemente cambia el modelo de negocio (DeNardis, 2007).

Un ejemplo claro es el conocido formato PDF, creado por una empresa comercial (cuyo objetivo principal es el beneficio económico) y que durante mucho tiempo fue propietario, por cuyo uso era preciso pagar licencia. En un momento determinado entendió que era preferible liberar el formato y centrar el negocio en una serie de productos que giraban en torno a él. De hecho la difusión de este formato provocó dos cuestiones, teniendo en cuenta que era usado por millones de personas debido, entre otras cosas, a que era abierto y libre, la venta de esos productos obviamente aumentó, pero además llegó un momento en que se convirtió en un estándar de derecho imponiéndose su uso para la conservación permanente de documentos de archivo (Une-ISO 19005-2008).

Internet, con la interconexión de ordenadores con circuitería, sistemas operativos y programas de aplicación muy diversos, ha sido un importante impulsor del uso de formatos abiertos. Los formatos del correo electrónico, o el HTML, el lenguaje de las páginas web son ejemplos claros de esto.

Un hito importante en el terreno de los formatos abiertos es la aparición del Open Document (OASIS), un formato ideado para documentos ofimáticos y diseñado por una alianza de numerosos desarrolladores y productores de software.

En este mismo terreno, es importante destacar los impulsos dados a uso de formatos abiertos por numerosas instituciones; unas veces bajo la forma de recomendaciones de uso; otras bajo formas más taxativas. Entre éstas, la reciente Ley 11 / 2007, de 22 de junio, de Acceso Electrónico de los Ciudadanos a los Servicios Públicos, prevé la configuración de la Administración electrónica y se inspira en varios principios. Uno de ellos es el denominado principio de neutralidad tecnológica, recogido en los artícu-

los 4. i) y 6. 2. c) y k), que significa sustancialmente que las Administraciones públicas deben asegurar que la elección tecnológica de los ciudadanos no obstaculice sus relaciones con la Administración y para ello deben utilizar también los denominados estándares abiertos.

En el terreno académico, son ya varias las universidades que adoptaron algún tipo de normativa o recomendación al respecto (Figuerola, 2007). En el mismo sentido la Conferencia de Rectores de las Universidades Españolas ha dictado unas Recomendaciones sobre intercambio de documentos en Universidades españolas mediante estándares abiertos (Conferencia de rectores, 2006).

3.2. Licencias copyleft

La concepción de la propiedad intelectual y derechos derivados en el entorno digital, y, sobre todo, la defensa de esos derechos requieren planteamientos distintos de los aplicados cuando se trabaja con información y soportes convencionales. La mera facilidad técnica para la copia, y las dificultades técnicas para evitarla, introducen elementos difíciles de tratar, e inducen a aplicar otro tipo de enfoques, en especial en lo que se refiere a modelos de negocio o explotación de los derechos derivados de la propiedad intelectual.

No es propósito nuestro entrar aquí en conocidas polémicas, sino hacer mención de iniciativas interesantes en este terreno. Una de ellas es lo que se dio en llamar copyleft, en contraposición con el concepto bien conocido de copyright (Laat, 2005). La idea básica es poner el acento en la libertad de uso y también de copia. Esto no quiere decir, como veremos a continuación, que esa libertad no tenga limitaciones, pero se estima que es más importante dicha libertad que las posibles restricciones (Figuerola, 2007). A este planteamiento se llega por diversas vías: desde la visión simplemente altruista hasta la que entiende que obtiene un mayor beneficio con la difusión cuanto mayor mejor de su obra; pasando por quienes piensan que, ya que les van a copiar, mejor ponerlo fácil y que lo hagan bien (por ejemplo, citando la autoría; respetando la integridad, etc.).

3.2.1. Creative Commons

Creative Commons es una fundación cuyo objetivo es traducir y adaptar a las legislaciones locales un conjunto de licencias de uso conocidas, precisamente, como Creative Commons o CC (Botterbusch, 2008). Las licencias CC, pues, son en realidad un conjunto de licencias que

surgen de combinar diversos elementos o atributos (Carbajo, 2007):

- atribución: obliga a citar las fuentes y autores de la obra licenciada
- no comercial: no permite que el uso de la obra licenciada tenga fines lucrativos
- compartir igual: obliga a que la obra sea distribuida sin cambios e íntegra
- no derivados: obliga a que las obras derivadas tengan siempre la misma licencia que la original

El autor puede aplicar a su obra cualesquiera de estos atributos, uno o varios (o todos). Las licencias CC se han revelado como un eficaz método para preservar la autoría de conocimiento cuyo medio preferente de circulación es Internet.

3.3. Open Archives Initiative (OAI)

Una de las iniciativas más importantes es aquella que busca facilitar la difusión de los contenidos digitales por medio de estándares de compatibilidad e interacción. Es lo que conoce como OAI (4). A través de ella cualquier institución puede hacer visibles sus contenidos en la red ya que sus sistemas no exigen el uso de programas propietarios y desde el punto de vista del diseño, permiten la incorporación de diferentes plataformas y paquetes, lo que se conoce con el nombre de "arquitectura abierta". Realmente lo más importante es la "interoperabilidad", que busca una mayor compatibilidad entre colecciones de contenido procedente de distintas fuentes. Gracias a esta iniciativa es posible localizar muchos contenidos en la web ya que esta iniciativa está presente tanto en los repositorios institucionales como en los disciplinares donde podemos localizar investigaciones, recursos para la docencia... La gran ventaja sin duda de estas alternativas es que no sólo se almacenan los documentos sino que cada documento lleva una serie de informaciones sobre el mismo que permiten su localización favoreciendo que las distintas instituciones puedan poner a disposición de la sociedad aquellos documentos que la institución ha ido generando o haciendo recopilaciones temáticas de cuestiones diversas. Ambas opciones, compatibles entre sí, favorecen también la conservación de los materiales.

3.4. Repositorios

Directamente en relación con el OAI están los llamados repositorios digitales. Éstos son, a muy grandes rasgos, almacenes donde diversas instituciones (especialmente universidades y

entidades relacionadas con la investigación científica) colocan los documentos, en sentido muy amplio, que producen (Lynch, 2003). Existen también repositorios temáticos, no necesariamente ligados a una institución determinada (Alonso, 2008).

Naturalmente, hablamos de almacenes en Internet, que toman la forma de portales web, y hablamos de documentos digitales. Muchas universidades disponen ya de su repositorio, y en ellos se cuelgan documentos de muy diverso tipo, desde tesis doctorales hasta artículos científicos, pasando por memorias de proyectos de investigación, informes técnicos, pre-prints ... En algunos casos se coloca allí información institucional, como estatutos, reglamentos, acuerdos de órganos de gobierno, y también el resultado de digitalizar documentos originalmente en papel u otros soportes; muchos de éstos tienen claro interés histórico.

En general, los repositorios tienen la característica de 'abiertos', es decir de libre acceso y, además, protegidos con licencias de tipo Creative Commons. Muchos de ellos, dependiendo de la institución de la que dependan, tienen mecanismos de autoarchivo; esto significa que los miembros de dicha institución son los que cuelgan en el repositorio los materiales de los que son productores. Este tipo de actividad se compatibiliza, en muchos casos, con el archivo en las mismas condiciones de tipo institucional.

Existen diversas aplicaciones informáticas diseñadas para implementar y gestionar repositorios, muchas de ellas de tipo software libre, aunque no todas. La mayor parte, o al menos las más usadas, permiten construir estructuras arborescentes, más o menos complejas, tendentes a reconstruir la estructura orgánica de la institución responsable del Repositorio. Todas siguen las especificaciones OAI, lo cual debería permitir que buscadores o recolectores especializados indizaran los documentos recogidos en el repositorio adecuadamente.

Sin embargo, OAI sólo cubre algunas necesidades. Ofrecido en dos vertientes, la más desarrollada, OAI-PMH (OAI-Protocol for Metadata Harvesting) tiene como objetivo el permitir esa recolección de registros por parte de motores de búsqueda especializados. La otra vertiente, OAI-Object Reuse and Exchange, tiene como misión definir estándares para la descripción e intercambio de agregados de recursos web. Los protocolos y estándares definidos por OAI son valiosos a la hora de exportar objetos o documentos digitales, en sentido amplio, pero distan mucho de haber resuelto algunos problemas importantes. Dejando de lado la multiplicidad de

esquemas descriptivos de dichos documentos, que ya en sí es un problema importante, OAI contempla los documentos como entidades individuales (que pueden, eso sí, estar formadas por varios objetos digitales), pero no recoge nada sobre la propia estructura institucional en la cual se inserta dicho documento.

Desde un punto de vista práctico, esto significa que la migración de repositorios a otras plataformas de software o a versiones diferentes de éstas se convierte en una operación complicada (Bell, 2006). Las plataformas informáticas que tenemos ahora no existirán siempre, y, de otro lado, los repositorios tienen interés en función, entre otras cosas, de su tamaño, de la cantidad de documentos digitales que contienen. Los protocolos OAI nos permiten, no sin dificultades, llevar documentos individuales de una plataforma a otra; pero cuando se trata de un repositorio completo tenemos que recolocar cada registro individual en su lugar de la estructura institucional que hemos debido reconstruir en la nueva plataforma. En repositorios que no son meros juguetes, esto significa miles de comunidades, subcomunidades, etc. y en cada una de ellas miles de documentos digitales, más o menos complejos.

3.5. URL persistentes

Lo que conocemos como URL (Uniform Resource Locator) es lo que actualmente utiliza Internet para identificar y localizar un objeto digital determinado, y esto incluye lo que denominamos documentos digitales en sentido amplio. El URL es la dirección de red que sirve para que nuestros programas obtengan en la red el objeto que sea.

Sin embargo, es dudoso que los URL puedan permanecer inalterados desde una perspectiva de tiempo media o larga. Los URL incluyen nombres de máquinas, dominios, subdominios, etc., además de rutas locales de directorios y ficheros. Con frecuencia, los recursos o documentos digitales cambian de máquina, éstas cambian de nombre (por ejemplo, porque son sustituidas por otras más recientes, los dominios y subdominios se reestructuran, etc. Incluso el concepto mismo y estructura de los URL es probable que no sea el mismo dentro de algunos años: no sabemos cómo será Internet en un futuro a medio plazo, ni siquiera si existirá como tal, o habrá otra cosa con otros esquemas de funcionamiento completamente diferentes (Deutsche Nationalbibliothek).

El resultado de todo esto es que los URL que nos servían para identificar y localizar un recurso digital cambian y se hacen inservibles. Cual-

quier planteamiento de preservación a medio plazo debiera prever esta contingencia. Hay diferentes soluciones propuestas, de las cuales citamos, a modo de ejemplo, alguna de las más conocidas.

3.5.1. *Persistent Uniform Resource Locator (PURL)*

PURL (OCLC) es en realidad un URL como cualquier otro, pero lo que hace es apuntar a un servicio intermedio que mapea o traduce el PURL al URL que actualmente apunta en realidad al documento o recurso que sea. Esto significa que, cuando la dirección o URL real de nuestro recurso cambia, por la razón que sea, hay que reflejar dichos cambios en el servicio que gestiona los PURLS; se supone que cada vez que creamos un recurso digno de ser preservado, obtenemos para él un PURL y hacemos que el servicio que gestiona dicho PURL lo vincule a la dirección real que en ese momento tiene el recurso en cuestión.

PURL es una idea auspiciada por OCLC, y, de hecho, OCLC gestiona un servidor PURL (3).

3.5.2. *Handle*

Aquí la dirección de cada recurso es identificada mediante un código asignado por un servicio dedicado a esta tarea (Corporation). La dirección o referencia de dicho recurso es, a todos los efectos, la dirección de dicho servicio más el código identificador. El servicio en cuestión localiza la dirección real del recurso y redirecciona de forma automática hacia él. Nuevamente, esto significa que hay que registrar cada recurso a preservar en dicho servicio, obtener un código y, naturalmente, comunicar a ese servicio cualquier alteración en la dirección o URL real del recurso de marras. Handle surgió del propio entorno de los desarrolladores de Internet y está definido formalmente en varias RFC. Utilizado por numerosas instituciones, es posiblemente el servicio de este tipo más difundido, tanto por su antigüedad como por su versatilidad (5).

3.5.3. *Digital Object Identifier (DOI)*

El DOI (6) es un sistema de identificación construido sobre el sistema Handle; básicamente es un código que se asigna a cada objeto que se desea registrar, que se compone de dos partes: una de ellas identifica a la entidad que ha otorgado el código, y la otra identifica al objeto individual en sí mismo. Esto implica la existencia de instituciones registradas que adjudican DOIs. Estas instituciones (Registration Agencies) están coordinadas por la DOI Foundation. Originariamente el uso más amplio del DOI provino

del sector editorial, aunque su uso se ha generalizado a todo tipo de objetos digitales.

El DOI tiene también otra vertiente interesante, relacionada con la protección de la propiedad intelectual de los autores de los objetos digitales registrados. Tal como se ha dicho en otro lugar de este trabajo, la facilidad de copia y difusión puede dar lugar a la proliferación de apócrifos y versiones derivadas. Independientemente de que esos derivados sean legales, permitidos por los autores originales, o no, el hecho de que una determinada versión esté identificada de forma unívoca (por ejemplo con un DOI) identifica también a sus autores.

Las Agencias que adjudican DOIs guardan el URL real de cada objeto, que debe ser actualizado cada vez que éste cambie. Pero guardan o almacenan también más información sobre él; este último aspecto puede ser altamente flexible y puede incluir, por ejemplo, un esquema de descripción basado en metadatos, como el famoso Dublin Core. Guardan también las peticiones donde se indica qué parte de esa información puede hacerse pública y en qué circunstancias.

4. Conclusiones

El predominio creciente de lo digital ofrece una serie de perspectivas imposibles e impensables antes, pero plantea también numerosos interrogantes, así como la necesidad de resolver determinados problemas. Las estructuras intelectuales que nos servían para ordenar el proceso documental convencional se muestran poco útiles a la hora de abordar situaciones no previstas, y el panorama es tanto más preocupante cuanto más lejos ponemos el horizonte temporal. Aplicamos soluciones parciales que funcionan tal vez ahora, pero de las que no sabemos el recorrido real que tendrán. La realidad va por delante y respondemos ante los nuevos retos a trompicones. En este trabajo hemos mostrado, sin ninguna intención de exhaustividad, algunos de los problemas que se nos plantean; y también algunas iniciativas interesantes para dar respuesta a esos problemas.

Y aunque esas iniciativas no están exentas de dificultades y sólo abordan aspectos parciales y concretos, queremos verlas desde el lado positivo, porque positivo debe ser el balance si contemplamos el panorama de posibilidades que nos ofrece el mundo digital. La existencia de aspectos problemáticos e incluso incontrolables en alguna medida, no es sino fruto de la eferescencia digital. El objetivo último de este trabajo no era otro que promover la reflexión y el debate sobre el impacto de lo digital en el mun-

do documental, con la vista puesta en el futuro; y, en ese sentido, esperamos haber contribuido en alguna manera.

Notas

- (1) Se pueden consultar algunos ejemplos de literatura colaborativa en las siguientes direcciones <http://www.calibreochoc.com>; <http://www.libroflotanete.net/>; <http://www.literativa.com/>.
- (2) Consultar las tablas 1 y 13 que aparecen en el Informe del PNUD de los años 2007-2008.
- (3) Se puede consultar más información en <http://purl.oclc.org>.
- (4) <http://www.openarchives.org/>.
- (5) Más información en http://en.wikipedia.org/wiki/Handle_System.
- (6) <http://www.doi.org/>.

Referencias

- Alonso, Julio, Subirats, Inma y Martínez Conde, M. Luisa (2008). Informe APEI sobre acceso abierto, Gijón, 2008.
- Asproth, V. (2005). Information technology challenges for long-term preservation of electronic information. *International Journal of Public Information Systems*. 1 (2005) 27-37. http://www.ijpis.net/issues/no1_2005/IJPIS_no1_2005_p3.pdf (2009-06-04).
- Bell, J. y Lewis, S. (2006). Using OAI-PMH and METS for exporting metadata and digital objects between repositories. // *Program: electronic library and information systems*. 40:3 (2006) 268-276.
- Borgman, C. (2000). *From Gutenberg to the global information infrastructure: access to information in the networked world*. MIT Press, 2000.
- Botterbusch, H. and Parker, P. (2008). Copyright and Collaborative Spaces: Open Licensing and Wikis. // *TechTrends*, 52:1 (2008) 7-9.
- Carbajo Cascón, F. (2007). Aproximación a las licencias Creative Commons. // *EnRed. Revista Digital de la Universidad de Salamanca*. 0 (2007) 9-11. http://enred.usal.es/hemeroteca/ENRED_0.pdf (2009-06-04).
- Collins-Sussman, B., Pilato, C., and Fitzpatrick, B. (2008). *Version control with subversion*. O'Reilly Media, Inc., 2008.
- Conferencia de Rectores de las Universidades Españolas. Comisión de TIC (2006). *Marco para el Intercambio de Documentos en Universidades Españolas mediante Estándares Abiertos*. 10 de julio de 2006 http://osluz.unizar.es/files/estandares_en_universidades.pdf (2009-06-04).
- Cordón García, José Antonio (2007). *La Europa de los ciudadanos: los retos del acceso y la conservación de la cultura*. // *El futuro de Europa: la Europa Social y de los ciudadanos*. Yuste: Fundación Academia Europea de Yuste, 2007.
- Corporation for National Research. *Handle System. Unique Persistent Identifiers for Internet Resources*. <http://www.handle.net> (2009-05-05).
- Dee Laat, P. (2005). Copyright or copyleft? An analysis of property regimes for software development. // *Research Policy*. 34:10 (2005) 1511-1532.
- DeNardis, L. and Tam, E. (2007). *Interoperability and Democracy: A Political Basis for Open Document Standards*

- 5th International Conference on Standardization and Innovation in Information Technology, Calgary (CA).
- Deutsche Nationalbibliothek. Persistent Identifier. Eindeutige Bezeichner für Digitale Inhalte. <http://www.persistent-identifier.de/?lang=en> (2009-05-05).
- Directiva 2001/29/CE del Parlamento Europeo y del Consejo, de 22 de mayo de 2001, relativa a la armonización de determinados aspectos de los derechos de autor y derechos afines a los derechos de autor en la sociedad de la información. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:ES:HTML> (2009-06-04).
- Figueroa, C., Berrocal, J., and Rodríguez, Á. (2007). El uso de software libre en los sitios web universitarios españoles. // *Actas del FLOSS International Conference*, celebrado en Jerez. <http://softwarelibre.uca.es/jornadas/fic/spa/materiales/actas.pdf> (2007-03-01).
- Figueroa, C. G. (2007). Open, Abierto ... // *EnRed. Revista Digital de la Universidad de Salamanca*. 0 (2007) 5-8 http://enred.usal.es/hemeroteca/ENRED_0.pdf (2009-06-04).
- Gantz, John F. and Chute, C. (2008). *El universo digital, diverso y en expansión acelerada. Un pronóstico actualizado del crecimiento de la información en el mundo hasta 2011*. International Data Corporation (IDC) and EMC. <http://chile.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf> 2009-06-04).
- Gómez Díaz, R. (2009). *Documentos y nuevas tecnologías*. // Sendín García, Miguel Ángel, Gómez Díaz, Raquel. Régimen jurídico de los documentos: aspectos administrativos, civiles, penales y procesales. Granada: Comares, 2009: 181-199.
- IFLA and IPA (International Publishers Association) (2002). *Preserving the Memory of the World in Perpetuity: A Joint Statement on the Archiving and Preserving of Digital Information*. <http://ifla.queenslibrary.org/V/press/ifla-ipa02.htm> (2009-06-04).
- Informe sobre el desarrollo humano 2007-2008: la lucha contra el cambio climático: solidaridad frente a un mundo dividido. Nueva York: Programa para las Naciones Unidas y el Desarrollo, 2007. ISBN: 978-84-8476-322-2 http://hdr.undp.org/en/media/hdr_20072008_sp_complet_e.pdf (2009-06-04).
- Kim, S., Myaeng, S., and Yoo, J. (2005). A Hybrid Information Retrieval Model Using Metadata and Text. // *Lecture notes in computer science*, 3815:232.
- Kuny, T. (1998). The digital dark ages? Challenges in the preservation of electronic information. // *International Preservation News*. 17 (1998).
- Ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos (BOE 23 de junio de 2007).
- Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? metrics for evaluating collaborative media as a news resource. // *Nature*. 2004.
- Lynch, C. A. (2003). *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*. // *IR Research*. <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>.
- Marcus, A. (2009). *Global/Intercultural User Interface Design*. Human-Computer Interaction: Design Issues, Solutions, and Applications, 2009.
- Mens, T., Wermelinger, M., Ducasse, S., Demeyer, S., Hirschfeld, R., and Jazayeri, M. (2005). Challenges in software evolution. // *Principles of Software Evolution, Eighth International Workshop on*, 2005. 13-22.

- OASIS. Oasis Open Document Format for Office Applications (OpenDocument) TC. <http://www.oasis-open.org> (2009-05-05).
- OCLC Online Computer Library Center. PURLS. A Project of OCLC Research.
- Rijsbergen, C. J. Van. (1979) Information Retrieval. London: Butterworths, 1979. <http://www.dcs.gla.ac.uk/Keith/Preface.html> (2009-06-04).
- The International DOI Foundation (IDF). The International DOI Foundation (IDF). The DOI System. <http://www.doi.org/> (2009-06-04).
- UNE-ISO 15489: 2006. Información y documentación. Madrid: AENOR, 2006.
- UNE-ISO 19005-1:2008 Gestión de documentos. Formato de fichero de documento electrónico para la conservación a largo plazo. Parte 1: Uso del PDF 1.4 (PDF/A-1).