

## [OpenAccessWeek Perú 2012](#)

**Miércoles 24 octubre 2012, 12.35-13.00 (videoconferencia)**

### **Preservación digital en cinco pasos**

Gracias a la organización de Open Access Week Perú por su invitación, un saludo fraternal a las profesoras Aurora de la Vega y Mónica Arakaki, que me han precedido, así como a John Hagen, Libio Huaroto, Raúl Sifuentes, Mónica Calderón, Sheila Blumen, y el resto de profesionales participantes en esta gran fiesta del acceso abierto.

Es un placer poder saludarles desde mi casa en Cataluña, España, y poder sumarme a Uds. gracias a las tecnologías, que espero nos respeten durante los próximos minutos.

Compartiré con Uds. algunas reflexiones sobre preservación digital, un aspecto del acceso abierto que no es menor si lo que nos planteamos desde nuestros servicios de información es garantizar el acceso permanente a la producción científica o cultural de nuestras organizaciones.

Sabemos que la preservación tradicional se ocupaba de mantener el documento en su estado original, incluso propiciando el aislamiento o el olvido de los documentos analógicos como forma de preservarlos para el futuro. El rico patrimonio peruano que todavía hoy sigue desconocido, y sin embargo protegido, es una buena muestra. Por contra, sabemos que la preservación digital lo es por alteración, y que el olvido es letal.

Por compartir con Uds. un ejemplo del lugar donde trabajo: la [Biblioteca de Cataluña](#), en Barcelona. Una de nuestras joyas patrimoniales es [Las Homilias de Orgañá](#), un sermulario en una incipiente lengua catalana de finales del siglo XIII, que fue hallado en 1904 en una pequeña iglesia del Pirineo catalán. ¿Por qué ha llegado hasta nuestros tiempos un documento del 1.200? Pues porque estuvo escondido, aislado del mundo, durante 700 años. Naturalmente dió la casualidad de que las condiciones climáticas y lumínicas fueron idóneas, pero ese aislamiento permitió su supervivencia.

Por contra, imaginemos que tomamos una página web actual, la alojamos en un servidor informático actual, y nos olvidamos de ella. En menos de diez años es probable que haya dejado de existir, bien porque el pago a la empresa que mantiene el servidor ha expirado, bien porque algunos de los componentes de esa página web ya no son legibles por nuestro navegador: sus imágenes o vídeos incrustados, sus diseños en flash, sus bases de datos dinámicas, sus hojas de estilo, etc.

¿Qué sabemos de preservación digital?

Sabemos que nuestros documentos digitales son frágiles y tienen una rápida obsolescencia.

Sabemos que después de multitud de estudios y proyectos no existe una solución técnica infalible que cumpla con todos los requerimientos de preservación digital. Actualmente se combinan estrategias a largo plazo y actividades experimentales para garantizar el acceso a los fondos digitales.

Sabemos que las dos técnicas más habituales, migración y emulación, deben ser empleadas lo mínimo posible, puesto que siempre nos van a traer complicaciones. La migración de un soporte a otro, de una cinta VHS a un DVD, y de un formato a otro, de Word a XML; y la emulación, más respetuosa pero poco experimentada, no son métodos infalibles.

Sabemos que debemos hallar el equilibrio entre la importancia relativa del documento y los factores de riesgo que pueden ser perjudiciales, puesto que los recursos son siempre limitados.

Sabemos también que a menudo debemos decidir entre preservar el contenido, incluso mejorando la legibilidad y utilidad gracias a las tecnologías existentes; o bien mantener el recurso preservando el aspecto y el comportamiento original, con las limitaciones que pueda ocasionar esta decisión.

Con estos precedentes, y dando respuesta a la invitación de la Pontificia Universidad Católica del Perú, he querido señalar cinco pasos que pueden ayudarnos en los procesos de preservación digital.

Uno, Crear una estrategia, una política de preservación digital. Dos, Confiar en OAIS como la mejor hoja de ruta para crear un sistema fiable de preservación digital. Tres, Centrarse en la ingesta, cuando ocurre la llegada del documento al sistema de preservación. Cuatro, Respetar las reglas en el uso de metadatos del documento. Y cinco, Cooperar con el resto de la comunidad para garantizar el acceso a nuestros propios avances en esta materia, la preservación digital, que está en una fase absolutamente prehistórica.

Vamos con los cinco pasos.

### PRIMER PASO. CREAR UNA ESTRATEGIA

Es decir analizar, pensar, tomar decisiones, y ponerlas por escrito.

En la planificación de una estrategia deberemos mantener un equilibrio entre los tres pilares, los tres factores determinantes de la preservación digital, que señalaron Kenney y McGovern, de The Library and Archives Canada, en 2005. Organización, tecnología, y recursos.

Organización: las políticas, culturas y procedimientos de cada biblioteca, archivo, o servicio de información, así como las responsabilidades de las personas y equipos implicados.

Tecnología: los avances técnicos existentes, que puedan ayudarnos a cumplir con nuestros objetivos, y el control de las amenazas de obsolescencia existentes.

Recursos: los medios económicos y humanos que van a hacer posible mantener el documento digital accesible a largo plazo.

Lo primero es tomar consciencia: la preservación digital es nuestra responsabilidad, no podemos esperar que sean otros los que vengan en el futuro a solucionarnos los problemas. Debemos tener una actitud proactiva y responsable, que marque nuestras decisiones.

Si vamos a planear un proceso de digitalización lo haremos pensando en una vez para todos los propósitos: preservación, acceso, difusión.

Si vamos a dar prioridad a determinados fondos o colecciones, por ejemplo los nacidos digitales, en detrimento de las colecciones híbridas o únicamente analógicas, es el momento de detallar el ritmo de las actuaciones.

Si vamos a caracterizar qué usuarios pueden depositar documentos en nuestro repositorio, y en qué condiciones deben hacerlo, es el momento de llevarlo a cabo, así como de plantear las probables cambios que esas decisiones van a comportar en el día a día de nuestras organizaciones.

Si tenemos que prever un presupuesto plurianual para garantizar que esos recursos sean estables en el tiempo, debemos ser capaces de priorizar las necesidades de nuestro sistema de preservación.

Las mismas autoras, Anne Kenney y Nancy McGovern, detallaron cinco fases de los programas de preservación digital. Ubicar nuestra organización en una fase u otra, o en un período u otro de esas fases, puede ayudarnos a planificar las siguientes etapas de nuestra infraestructura y cultura de preservación digital.

La primera se refiere justamente a la identificación de la fase en la que se halla nuestra organización: ¿tal vez estamos negando que la preservación digital puede ser un problema? La segunda sugiere el paso a la acción, solucionando pequeños problemas, o definiendo proyectos que persiguen evitar dificultades concretas. La tercera fase establece la consolidación de un programa de preservación, y la cuarta la integración de las acciones de preservación en los flujos de trabajo de nuestra organización. La quinta, fase superior de este modelo, implica la participación de toda la organización en la estrategia y la táctica de preservación digital.

Identificar la situación de nuestra organización y detallar los pasos a seguir para elevar el compromiso de la misma con los procesos de preservación digital es un paso indispensable antes de llevar a cabo el diseño de un flujo de trabajo dedicado a la preservación digital.

## SEGUNDO PASO. CONFIAR EN OAIS

OAIS (Open Archival Information System) es una norma ISO, la 14.721 que cuenta con diversas [ediciones](#) e informes precedentes. Todos Uds. tienen acceso a la documentación que describe el sistema OAIS, si bien debemos reconocer que no siempre esa información es suficientemente comprensible para usuarios no especializados.

Desarrollado inicialmente por la NASA, intenta crear una regla donde no las había. Una hoja de ruta con dos funciones básicas: preservar y proveer acceso a la información.

Su objetivo básico: conseguir que la información llegue al futuro y sea comprensible por parte de los futuros usuarios.

Un ejemplo recurrente para explicar cómo funciona OAIS es el de Gladney, de 2006, que fue recogido por Alice Keefer y Núria Gallart en su [libro](#) sobre preservación digital, uno de los pocos manuales existentes en español. Si OAIS quisiera explicar cómo es un automóvil, nos explicaría que es un vehículo de cuatro ruedas, capaz de transportar entre dos y ocho personas a una velocidad posible de 130 km/h y un espacio para equipaje y con prestaciones para que el viaje sea más agradable, como reproductor musical, aire climatizado, asientos cómodos, etc.

Básicamente, OAIS establece criterios para determinar qué materiales deberán depositarse en un sistema de archivo fiable, y en el mismo momento intenta motivar a los productores para transferir adecuadamente esos documentos al repositorio seguro.

Se asignan derechos de depósito, transformación y difusión, para cada perfil de usuario, y se estimula la creación de políticas y documentación de procesos.

En OAIIS se prevee el tratamiento en paquetes que contienen por una parte el objeto a preservar, y por otra los metadatos. Ese paquete de seguridad sufre alteraciones según sea el momento de la ingesta, de la preservación, o de la difusión a los usuarios autorizados. El momento más importante, como veremos, es la ingesta, pues de las decisiones y precauciones iniciales va a depender la supervivencia del documento.

### TERCER PASO. CENTRARSE EN LA INGESTA.

El 90% de la preservación digital ocurre en este momento del proceso, cuando el objeto depositado, junto a sus metadatos, se convierte en el objeto a preservar.

Aunque parezca una obviedad, mi recomendación primera es que consigan un buen antivirus y sometan al documento depositado a un análisis en profundidad, antes de permitir una epidemia en su sistema.

En este momento deberán también conseguir que los propietarios de los derechos del documento les autoricen, como mínimo, a tres cosas. Una, al depósito que están a punto de realizar. Dos, a transformar en el futuro el documento, si ello es necesario para su preservación. Y tres, a dar difusión al documento y en qué condiciones, sea por medio de la comunicación pública, o por medio del acceso restringido a un perfil determinado de usuarios. El inicial es el momento de conseguir las autorizaciones del futuro.

Otra medida inicial esencial es la elección del formato del objeto digital, así como la identificación de los metadatos que vamos a requerir en el proceso de depósito.

Según la NARA existen unos 16.000 tipos de formatos. Desde los clásicos TXT en ASCII, TIFF o WAV, a los ya desaparecidos Wordperfect o Wordstart, por poner dos ejemplos clásicos de procesadores de texto.

En los procesos de ingesta seremos restrictivos, puesto que queremos contribuir en ese momento inicial a su posterior recuperación. Daremos prioridad a los formatos estándares de código abierto, los recomendados por las instituciones de referencia, y aceptaremos con reservas los formatos propietarios que tengan una base de clientes muy amplia, puesto que confiamos en la fuerza de la comunidad internacional para hallar soluciones a los problemas de acceso que puedan existir. Limitaremos al máximo la entrada en nuestro sistema de formatos propietarios de difusión limitada, puesto que en el futuro van a significar, muy probablemente, problemas para nuestro sistema.

Por poner un ejemplo, aceptar en nuestro sistema un mensaje de correo informático proporcionado por un sistema de gestión de correo basado en un viejo software propietario de difusión limitada nos va a traer muchos problemas. Aceptarlo en un formato creado por Outlook Express no es recomendable, pero sí aceptable. Emplear XML para los mensajes de correo electrónico es lo mejor, puesto que ese estándar de código abierto va a ser legible en el futuro sin demasiadas complicaciones.

La elección no está exenta de dificultades: imaginemos que queremos conservar las pistas de audio de nuestra discoteca. Por un lado tenemos ACC (Advanced Audio Coding), el formato propietario al que iTunes pasa las canciones adquiridas por dispositivos Apple, claramente no recomendable. Por otra parte, tenemos MP3, también un formato propietario con bajo nivel de compresión y amplia base de clientes, creado a partir de diversas patentes, algunas de las cuales han sido ya publicadas en abierto. MP3 es aceptable. Finalmente, tenemos el formato WAV, creado por Microsoft, que es el formato preferido, aunque no existe consenso en las recomendaciones internacionales, y por ello debemos estar al corriente de su evolución.

En la fase de ingesta deberemos aplicar a los documentos depositados las primeras acciones de migración de soportes o actualización de formatos, si ello es imprescindible para su recuperación futura.

#### CUARTO PASO. RESPETAR LAS REGLAS.

El éxito de nuestros sistemas de información se basa en garantizar la interoperabilidad de los sistemas. La existencia de normas y reglas de descripción han hecho posible el intercambio de información, y contribuirán a la supervivencia de los documentos digitales.

El uso de metadatos es esencial para que los sistemas basados en OAIIS funcionen. Los metadatos deben permitir que una persona sin conocimientos previos sobre el formato del documento digital pueda acceder a él y leerlo sin problemas.

Existen tres tipos de metadatos, si bien el cumplimiento estricto de esos esquemas descriptivos puede quedar a criterio de la organización, dependiendo del tipo de información que gestiona, y sus necesidades futuras.

En primer lugar, los metadatos descriptivos: los que describen el contenido del documento y le adjudican componentes semánticos: qué contiene y sobre qué, cuál es el título y la autoría, en qué fecha se creó y qué palabras clave pueden definirlo. Las normas MARC21, ISAD(G) o Dublin Core están especialmente recomendadas.

En segundo lugar, los metadatos técnicos: los que describen el continente del documento, le adjudican derechos de acceso, y determinan qué pasos se han realizado para su preservación. Los técnicos describen con qué sistema operativo fue creado, qué tamaño y formato tiene el fichero. De entre los metadatos técnicos cabe destacar por una parte los referidos a derechos de acceso, que especifican la autoría, la propiedad de los derechos y el copyright; y por otra los metadatos de preservación, que narran el historial de acciones que se han realizado para preservarlo: las potenciales migraciones de un formato a otro, la copia de los ficheros o su transformación de un sistema operativo a otro, los resultados de los antivirus a lo que ha sido sometido, etc. Los esquemas METS, para los datos puramente técnicos, y PREMIS, para los de preservación, fueron creados para dar respuesta normalizada a estas necesidades.

En tercer lugar, los metadatos estructurales: aquellos que establecen relaciones entre el objeto descrito y otros objetos contenidos en el mismo repositorio: los capítulos de un libro, un e-mail con sus anexos, las diversas versiones de una grabación musical, etc.

Existen herramientas, que aunque precarias, nos pueden ayudar: el paquete FITS, por ejemplo, contiene diversas aplicaciones informáticas, como JHOVE, DROID (que bebe de la base de datos PRONOM), o el metadata extractor de la National Library of New Zealand, que pueden ayudarnos a extraer metadatos técnicos, como los referidos al tipo de fichero que es nuestro documento, y sus características. Pero para ello, para que la tecnología pueda leer estos sistemas, debemos respetar las reglas de descripción, y utilizar esquemas y normas estándares.

## QUINTO PASO. COOPERAR

Hablar de cooperación en la semana del Open Access puede parecer redundante, pero en preservación digital es básico recurrir al lema de los mosqueteros, “Uno para todos, todos para uno”.

Muchas de las iniciativas emergentes van en la lógica de compartir recursos y cooperar para contribuir al éxito de las estrategias de preservación digital. La duplicación del contenido de los servidores destinados a preservación es una medida de seguridad esencial. Si algo ocurre a nuestro servidor, sea un virus o sea una catástrofe natural, sea un incendio o un descuido sin intención, contar con réplicas de nuestras colecciones en otros lugares, alejados geográficamente, puede servir para subsanar el problema.

La herramienta LOCKSS, por ejemplo, no solo colecciona y conserva en los cachés de bibliotecas de todo el mundo los fondos de los repositorios distribuidos, sino que

lleva a cabo rutinas de trabajo propias de los sistemas de preservación --analizando las cadenas de bits, esencialmente--, para garantizar que los datos allí preservados se mantienen tal y como se depositaron. Por supuesto, el software de LOCKSS y la participación en su red son gratuitos.

[Lluís Anglada](#) es uno de los expertos más solventes en materia de cooperación bibliotecaria, y actualmente director del Consorcio de Bibliotecas Universitarias de Cataluña. Cuando se refiere al nacimiento en 1996 del consorcio que dirige lo hace con estas palabras: el nacimiento del consorcio es fruto de diversas cosas, de la oportunidad de la amistad, del momento histórico, pero sobre todo es fruto de la voluntad de hacer. De tener en mente la idea de que para mejorar no es suficiente con mejorar la casa de cada uno, sino que debemos mejorar también el edificio donde vivimos y tal vez, si podemos, la calle donde está ese edificio.

Esa es la cooperación necesaria para que la preservación digital sea un éxito.

Recuerden: creen su propia estrategia, confíen en OAIS, céntrense en la ingesta, respeten las reglas y normas existentes, y cooperen con el resto de la comunidad para garantizar el acceso a su patrimonio digital.

Muchas gracias por su atención y por la oportunidad de participar en la Open Access Week del Perú. Estoy a su disposición.

[Ciro Lluca](#) es un bibliotecario y profesor universitario experto en preservación digital.

Coordinador del Patrimonio Digital de Cataluña (PADICAT) en la Biblioteca de Cataluña, España.

Profesor asociado en la Universidad de Barcelona (Grado de Información y Documentación; Máster de Gestión Cultural; Máster de Bibliotecas y Colecciones Patrimoniales). Profesor colaborador en la Universidad Pompeu Fabra (Máster online en Documentación Digital: módulos de “repositorios digitales”, y “archivo de Internet”).

Editor para España y coordinador editorial para Europa del repositorio E-LIS, *Eprints in Library and Information Science*.

Diplomado en Biblioteconomía y Documentación por la Universidad de Barcelona. Licenciado en Documentación y Posgraduado en Servicios de Información en las Organizaciones por la Universitat Oberta de Catalunya (UOC). Máster en Documentación Digital por la Universidad

Pompeu Fabra. Especializado en Comisariado y conservación de arte digital (2012) por la UOC. *Digital Preservation Training Programme* (2012) por la University of London Computer Center.

Autor de diversos artículos y ponencias sobre repositorios digitales, preservación digital y el archivo de Internet. Ha impartido charlas profesionales y sesiones formativas en España (recientemente: coordinador del seminario “El Archivo de Internet: bibliotecas que piensan en el futuro”, en 12as Jornadas Españolas de Documentación - Fesabid 2011; instructor del curso “Repositorios digitales” junto a Sandra Reoyo, en COBDC 2012); México (instructor del curso “Repositorios digitales”, en UNAM 2011); y Perú (formador del taller “Preservación digital” y el curso “Repositorios digitales”, en UNMSM 2012; así como charlas profesionales, en PUCP 2012, y en ETD Lima 2012: Simposio Internacional de Tesis y Disertaciones Electrónicas).

Fue vicepresidente del COBDC, Colegio Oficial de Bibliotecarios-Documentalistas de Cataluña (2003-2006).