

Soportes y formatos de conservación y transferencia: Hacer elecciones responsables

Jim Suderman

Presentado en la Conferencia SARBICA, Hanoi, 5 de Mayo de 2004

Fuente: Cedido por el autor a partir del documento original en InterPARES 2: <http://www.interpares.org>

Traducción, Alejandro Delgado Gómez

La elección de soportes y formatos de conservación y transferencia es un tema práctico, y como tal cualquier discusión acerca de él debe tener lugar dentro del contexto del conservador específico –particularmente en relación con los recursos disponibles para el conservador. No está particularmente regulado por principios teóricos. La primera parte de este ensayo revisará los soportes elegidos para transferencia y conservación de documentos electrónicos en instituciones líderes de Canadá, los Estados Unidos y el Reino Unido¹. La segunda parte revisará los formatos de transferencia y conservación. Esencialmente, estas secciones son simples listas. La última sección esbozará brevemente algunas consideraciones o recomendaciones para elegir soportes y formatos de conservación. Es aquí donde se abordará el elemento “Hacer elecciones responsables” del título de la sesión.

Soportes de transferencia y conservación

File Transfer Protocol (FTP) – actualmente los ficheros “ no clasificados, no comprimidos, no restringidos y no encriptados” son los únicos aceptados vía FTP en NARA². Library and Archives Canada está considerando una recomendación para utilizar FTP y el Protocolo Web Distributed Authoring and Versioning (WebDAV) como mecanismos de transferencia en línea³.

Soportes en cinta (transferencia o almacenamiento)⁴

- cinta de cartucho Cinta Lineal Digital (DLT IV) (no comprimida);
- cinta de cartucho Cinta Lineal – Abierta (LTO);

¹ Por favor, nótese que esta información es considerada exacta por el autor en el momento de la presentación, pero no debiera confiarse en ella para la toma de decisiones. Las instituciones que consideren la selección de soportes de conservación o transferencia debiera hacer consultas directas a otros depósitos de documentos electrónicos.

² “Expanding Transfer Options for Electronic Records” en el sitio web de NARA en http://www.archives.gov/records_management/initiatives/transfer_regulation.html, consultado el 18 de marzo de 2004.

³ Mike Swan. “Library and Archives of Canada: Electronic File Transfer Analysis” (2003), pp. 4-5. Este informe también examina las medidas de seguridad y valora las aplicaciones de interfaz de usuario.

⁴ Este listado incluye todos los soportes utilizados para almacenamiento o transferencia de documentos electrónicos en los National Archives de los Estados Unidos, los National Archives del Reino Unido, y Library and Archives Canada.

- cinta de 8 mm⁵;
- cinta de carrete abierto de 9 pistas;
- cartuchos de cinta clase 3480;
- Disco Compacto – Grabable (CD-R);
- “Disco giratorio” (almacenamiento en línea o casi en línea)

Este último soporte de almacenamiento está actualmente en uso en algunas instituciones, incluidos los National Archives of Australia, creo. En mi institución el almacenamiento en disco ha sucedido a los soportes en cinta, y estuvo causado por la consciencia de que no teníamos las destrezas necesarias para almacenar documentos en soportes en cinta de manera eficaz y fiable. Combinada con ello existía una actualización de nuestra tecnología base, de Windows NT a Windows 2000. Las cintas de conservación se habían realizado en entorno NT y eran inaccesibles en el entorno 2000, siendo necesaria una resolución inmediata.

Formatos de transferencia y conservación

En 2002 los National Archives del Reino Unido iniciaron un notable nuevo proyecto llamado PRONOM. El Proyecto PRONOM reúne y hace disponible información técnica acerca de estructuras de formatos de ficheros y productos de software que los soportan.

Esto se ve como un prerrequisito para cualquier régimen de conservación digital. Esta aproximación a la conservación representa una aproximación a la conservación significativamente diferente de la propuesta por el documento verde de los National Archives de Australia (al que se hizo referencia en una sesión anterior). Es similar en el hecho de que representa un compromiso institucional significativo y permanente para retener conocimiento y especificaciones particulares como parte del programa de conservación. Parte de este compromiso es establecer y mantener contactos con los principales desarrolladores de software para información de los productos de software. Ambas aproximaciones parecen adoptar el principio de que gran parte de la inversión en conservación debiera realizarse desde el comienzo, más que confiar en futuros desarrollos de la tecnología para abordar los problemas de la conservación. Los planes de los años venideros verían un PRONOM mejorado para incluir “herramientas específicas para apoyar actividades de conservación digital”⁶.

⁵ Library and Archives Canada invirtió inicialmente en el disco óptico STORLORD como soporte de conservación. Este soporte es ahora obsoleto y los fondos se están transfiriendo a cinta de 8 mm. Puede que este proceso se haya completado en el momento de esta presentación.

⁶ Véase “Background” en el sitio web de PRONOM en <http://www.records.pro.gov.uk/pronom/background.htm> (consultado el 18 de marzo de 2004).

El siguiente listado de formatos de transferencia procede de Library and Archives Canada⁷. Como PRONOM, esta iniciativa representa una inversión institucional significativa en la medida en que los formatos identificados en esta recomendación están sujetos a las rápidas transiciones de los desarrollos tecnológicos y cambios en el mercado. Por esta razón está programada la revisión anual de la recomendación, con actualizaciones a medida que sean necesarias. Los motivos para la selección de estos formatos son tan aplicables desde el punto de vista de la conservación como desde el punto de vista de facilitar una transferencia de documentos.

Aunque Library and Archives Canada tiene la capacidad tecnológica para tratar todos los formatos de fichero identificados, los formatos son o "recomendados" o "aceptables".

Los formatos recomendados son aquellos que soportan transferencias tanto de documentos a los archivos, como entre agencias gubernamentales, esto es, mientras los documentos todavía están activos. Los formatos aceptables son aquellos que sólo debieran utilizarse para la transferencia de documentos a los archivos. En su selección de formatos de ficheros, se equilibraron los requisitos de calidad, estabilidad, longevidad del documento y su aceptación por la industria. Nótese que la distinción entre formatos recomendados y aceptables se basa puramente en motivos tecnológicos, como opuestos a procesos de negocio⁸.

Series de caracteres

- American Standard Code for Information Interchange (ASCII) [ISO/IEC 8859-1:1998 (Latin-1)]
- Extended Binary Coded Decimal Interchange Code (EBCDIC)
- Unicode Version 3.0 UTF-8 [ISO/IEC 10646-1:2000]

En la actualidad NARA acepta sólo series de caracteres ASCII y EBCDIC, aunque está en progreso la consideración de otros formatos⁹.

Audio digital

- Audio Interchange File Format (AIFF)
- WAVE (WAV)
- MPEG-1: Layer 3 (MP3)

⁷ Electronic Records Development Division, Library and Archives Canada. "Guidelines for Computer File Types, Interchange Formats and Information Standards" (Version 1.0, February 2004). Los National Archives [del RU] "no especifican ni adoptan el uso de ningún formato de fichero particular." Véase Adrian Brown, *Digital Preservation Guidance Note 1*, "Selecting File Formats for Long-Term Preservation," p. 4, disponible en <http://www.pro.gov.uk/about/preservation/digital/guidance/default.htm>, consultado el 22 de marzo de 2004.

⁸ De la introducción a "Guidelines for Computer File Types, Interchange Formats and Informations Standards", p. 7. Para consultas concernientes a la asistencia específica acerca de qué formatos se recomiendan, por favor contacte a Electronic Records Development Division, Library and Archives Canada, 344 Wellington Street, Ottawa, Canada, K1A 0N3 (email: imgi@archives.ca)

⁹ Véase NARA Regulations en el Code of Federal Regulations, § 1228.270 Electronic records en http://www.archives.gov/about_us/regulations/part_1228_l.html, consultado el 22 de marzo de 2004.

- Musical Instrument Digital Interface (MIDI)
- Real Audio (RM/RA)

Imágenes digitales

- International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) T.4 y T.6
- Portable Network Graphics (PNG)
- Tagged Image File Format (TIFF)
- Graphics Interchange Format (GIF)
- Joint Photographic Experts Group (JPEG) [ISO/IEC 10918-1:1994]
- JPEG File Interchange Format (JFIF)

NARA acepta imágenes formateadas en JPEG/JFIF y TIFF, sujetas a ciertas especificaciones de calidad de imagen¹⁰.

Video digital

- Moving Pictures Expert Group (MPEG-2)
- Audio Video Interleave (AVI)
- MPEG-4
- Quicktime (MOV)
- RealVideo de Real Networks (RM)

Documentos - Textuales

- Extensible Markup Language (XML)
- Extensible HyperText Markup Language (XHTML)
- HyperText Markup Language (HTML)
- Standard Generalized Markup Language (SGML) [ISO/IEC 8879:1986]
- Ficheros de texto (.txt)
- Formato de documento Microsoft Word (.doc)
- Portable Document Format (PDF)
- Formato de documento WordPerfect (.wpd)

Los documentos textuales existen con creciente frecuencia en formatos de imagen escaneada. NARA aceptará imágenes escaneadas de documentos textuales en formatos TIFF, GIF, BIIF (Basic Image Interchange Format), y PNG, sujetas a ciertas especificaciones de calidad¹¹.

¹⁰ Véase "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records - Digital Photographic Records" en http://www.archives.gov/records_management/initiatives/digital_photo_records.html, (consultado el 22 de marzo de 2004).

De manera similar, NARA aceptará documentos creados y convertidos a formato PDF, sujetos a ciertos requisitos, por ejemplo la retirada o desactivación de la configuración de seguridad¹².

Correo electrónico

- Extensiones de correo de Internet multipropósito (MIME)

NARA aceptará correos electrónicos en lenguaje de marcado normalizado (preferiblemente) o formatos nativos, ambos sujetos a ciertas especificaciones. El formato PDF se identifica explícitamente como un formato inaceptable para correos electrónicos, aunque ese formato puede ser aceptable para adjuntos¹³.

Datos geoespaciales

- Digital Line Graphs – Level 3 (DLG-3)
- Environmental Systems Research Institute (ESRI) Export Format – (E00)
- Environmental Systems Research Institute (ESRI) Shape Format – (SHP)
- GeoTIFF
- Geography Markup Language (GML), Version 3
- International Hydrographic Organization (IHO) S-57, Edition 3.1
- TC 211 ISO 191xx Standards for Geographic Information
- Spatial Data Transfer Standard (SDTS)
- Canadian Council on Geomatics Interchange Format (CCOGIF)
- CARIS ASCII
- CEOS Superstructure Format
- Digital Elevation Model (DEM)
- GeoVRML (Virtual Reality Modelling Language)

Datos estructurados – Bases de datos y hojas de cálculo

- Ficheros planos
- Formato dBase (DBF)

Diseños técnicos

¹¹ "Expanding Acceptable Transfer Requirements: Transfer Instructions for Existing Permanent Electronic Records – Scanned Images of Textual Records" en http://www.archives.gov/records_management/initiatives/scanned_textual.html, (consultado el 22 de marzo de 2004).

¹² "Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records – Records In Portable Document Format (PDF)" en http://www.archives.gov/records_management/initiatives/pdf_records.html, (consultado el 22 de marzo de 2004).

¹³ "Expanding Acceptable Transfer Requirements: Transfer Instructions for Existing Email Messages with Attachments" en http://www.archives.gov/records_management/initiatives/email_attachments.html, (consultado el 22 de marzo de 2004).

- Drawing Interchange File Format (DXF)

Hacer elecciones responsables

En último extremo la elección de los formatos y soportes de conservación debe satisfacer las responsabilidades del programa archivístico: la conservación segura y la reproducción fiable de documentos electrónicos. Esas responsabilidades se satisfacen a través de elecciones que se basan en las tecnologías que nos son accesibles y a las que nuestras instituciones pueden hacer frente; a través del conocimiento del efecto de esas tecnologías sobre los documentos y el acceso a ellos. El documento verde australiano citado previamente sugiere que "gran parte de los esfuerzos de conservación tienen que invertirse al comienzo, no en el mantenimiento de un emulador continuo ni en conversiones de datos"¹⁴. Entiendo que esto significa que parte de la responsabilidad del programa de conservación es no sólo estabilizar documentos electrónicos para su conservación, sino también estabilizar el coste permanente de mantenerlos dentro del depósito archivístico.

El ritmo del cambio tecnológico significa que cualquier selección de soportes o formatos de conservación es temporal. Puesto que no existe ningún énfasis archivístico particular sobre el soporte, hacer una elección responsable aquí se basa primariamente en factores que afectan a la institución. Puesto que el esfuerzo archivístico no trata sólo de conservar documentos, quizá la consideración archivística primaria a tener en cuenta cuando se selecciona un soporte de conservación es su impacto en la provisión de acceso a los documentos.

Los National Archives del Reino Unido identifican el almacenamiento en disco duro basado en un servidor como "el régimen de almacenamiento de documentos electrónicos más eficaz y seguro, supuesto que esté bien gestionado e incluya una estrategia de copia de seguridad eficaz"¹⁵. Se reconoce que puede que tal almacenamiento no siempre sea posible e identifica seis criterios para valorar los soportes de almacenamiento eliminables. Estos son:

- Un ciclo de vida probado de al menos diez años;
- Una capacidad de almacenamiento adecuada a la cantidad de datos a ser almacenados;
- Una alta calidad en la detección y corrección de errores;
- Una tecnología madura con una presencia en el mercado bien establecida;

¹⁴ Helen Heslop, Simon Davis, Andrew Wilson. "An Approach to the Preservation of Digital Records" (National Archives of Australia Green Paper, 2002), p. 13. Disponible en http://www.naa.gov.au/recordkeeping/er/digital_preservation/Green_Paper.pdf, consultado el 22 de marzo de 2004.

¹⁵ Adrian Brown. "Selecting Storage Media for Long-Term Preservation" The National Archives. *Digital Preservation Guidance Note: 2* (2003), p. 4. (Disponible en <http://www.pro.gov.uk/about/preservation/digital/guidance/default.htm>, consultado el 22 de marzo de 2004.)

- Una buena recuperación de la inversión (incluido el precio por MB y los costes totales de comprar y mantener el hardware y el software necesarios);
- Ser durable tanto en términos de las cubiertas o capas protectoras como, para soportes magnéticos, con un alto valor coercitivo¹⁶.

A estos útiles criterios podrían añadirse

- Costes de nueva formación para el personal responsable de gestionar los soportes de almacenamiento;
- Costes de espacio de almacenamiento, por ejemplo, estanterías convertibles para acomodar diferentes factores de la forma del soporte;
- La posibilidad de que quizá haya que cambiar los actuales procesos de trabajo – nótese que los nuevos soportes pueden afectar a los procesos de trabajo tanto para la conservación como para la provisión de acceso a los documentos

El requisito de base de InterPARES de que la conservación de documentos electrónicos auténticos requiere conocer el modo en que los documentos serán afectados cuando se reproduzcan convierte a la selección de formatos en algo más retador. Los formatos que son satisfactorios para otros fines puede que no reproduzcan el documento de manera satisfactoria. Los formatos de conservación satisfactorios también pueden complicar la provisión de acceso a los documentos. Por ejemplo, las transferencias de correo electrónico desde muchas cuentas diferentes de correo electrónico. Puede ser más simple transferir todos los correos electrónicos a un formato de base de datos. Si se adquirió toda la base de datos de correos electrónicos, puede que sea difícil limitar el acceso a documentos individuales de correo electrónico, o a los correos de un solo usuario. Si se adquieren los ficheros individuales que reflejan los correos electrónicos de cada usuario individual, puede que sea difícil acceder a todo el cuerpo de correos electrónicos. La elección del formato de conservación afectará al modo en que puede proporcionarse acceso a los documentos. Puesto que este tipo de valoración puede ser intensa, el cambio de formato de los documentos debiera realizarse de manera tan infrecuente como fuera posible, y la longevidad del formato elegido debiera ser una consideración importante en este proceso¹⁷.

La elección de los soportes de conservación está dominada por el mercado, en términos de qué soportes están disponibles, y el coste al que el conservador puede hacer frente. Estos factores también son relevantes cuando se eligen formatos de conservación. En este último caso, sin embargo, el impacto del formato sobre la capacidad para reproducir de manera fiable

¹⁶ Se recomienda que estén por encima de los mil oersteds.

¹⁷ "Para disminuir el riesgo a la integridad de los documentos, el programa de conservación minimizará el número de tratamientos de conservación aplicados a todos los documentos digitales." Heslop, Davis, Wilson. "An Approach", p. 16.

documentos auténticos es primordial. Por estos motivos ambas consideraciones son relevantes en términos de la gestión de la evaluación archivística y los procesos de conservación.