Políticas de transferencia de documentos electrónicos

Jim Suderman

Fuente: presentado a la Conferencia SARBICA, Hanoi, 3 de mayo de 2004

Traducción: Alejandro Delgado Gómez

El Modelo de Conservación de InterPARES 1 identifica cinco factores que afectan a las transferencias de documentos electrónicos desde un creador o usuario a un conservador¹. Estos factores son:

- la política de ingreso;
- la infraestructura tecnológica del conservador (incluidos los conjuntos de habilidades disponibles);
- una estrategia de conservación para un cuerpo particular de documentos;
- un método objetivo de conservación (i.e., el modo en que se ejecutará la estrategia); y
- los componentes digitales reales y la información complementaria de los documentos transferidos.

Una política de ingreso se define como la política del conservador para aceptar la responsabilidad de los documentos transferidos. Esta política identifica normas y especificaciones para los documentos transferidos. Así, podría abordar consideraciones como

- formatos aprobados de transferencia de ficheros (p. ej., documentos de base de datos a ser transferidos en formatos de valor separado por coma);
- qué debiera hacerse cuándo los documentos se pierden o está incompletos; o la documentación se pierde o es poco clara;
- una fecha de transición que establezca el momento en que las responsabilidades de provisión de acceso a los documentos se desplazan del creador al conservador.

Con un programa de documentos electrónicos en su infancia, la política de ingreso de mi institución se enfoca primariamente en torno a la comprobación de las transferencias por

¹Los cinco factores se refieren a las flechas que apuntan a la actividad "Traer documentos electrónicos" (A2) identificada en el modelo "Conservar documentos electrónicos". InterPARES Preservation Task Force. *Model diagrams*, disponible en http://www.interpares.org/book/interpares_book_n_app05ii.pdf, chequeado el 29 de marzo de 2004. Las definiciones de los factores proceden de Preservation Task Force. *Arrow definitions*, disponible en http://www.interpares.org/book/interpares_book_n_app05iv.pdf, chequeado el 29 de marzo de 2004.

adelantado. Puesto que no se ha desarrollado ningún equipamiento para la transferencia en línea de documentos, los documentos electrónicos se transfieren sobre soportes digitales, como cinta o disco óptico. La escritura de datos sobre cinta se realiza usualmente utilizando software de copia de seguridad que es altamente propietario, así que antes de la transferencia real se negocian los soportes y los formatos. La configuración acordada se prueba posteriormente y, si tiene éxito, se utiliza como la base para la transferencia real de documentos. Al mismo tiempo, se confirman los formatos de fichero elegidos. Un listado de formatos de fichero aceptables aún no ha sido finalizado. Muchos de los ficheros ofimáticos se aceptan en su formato nativo, i.e., en el mismo formato con el que fueron creados o utilizados. El formato de fichero preferido para imágenes es formato de fichero de imagen etiquetado (.tiff), pero también pueden aceptarse el del grupo conjunto de expertos en fotografía (.jpeg) y el formato de intercambio de gráficos (.gif). El formato de base de datos (.dbf) es el formato preferido para bases de datos. El formato de valor separado por coma (.csv) puede ser aceptable, especialmente para aplicaciones de hoja de cálculo. Estos formatos han sido aceptados por un cierto número de archivos en Canadá.

Después de la transferencia real, se emprenden comprobaciones adicionales, que comienzan con una confirmación básica de que los documentos recibidos son, de hecho, los documentos que se esperaban. Siempre que es posible también confirmamos el número total de documentos. En un caso hubo una discrepancia de unos 10.000 ficheros respecto de lo que esperábamos. Comunicamos esta discrepancia al cuerpo que los transfirió y resolvimos el asunto. Esto es importante para confirmar que hemos recibido todo lo que tenemos que haber recibido para conservar un documento completo.

Además de estas consideraciones tecnológicas, usualmente existe un protocolo establecido para regular las peticiones de acceso a los documentos. En Ontario éstas están reguladas por la ley *Freedom of Information and Protection of Privacy*. Este protocolo también regularía las peticiones de recuperación de documentos legales. Normalmente las solicitudes de información o de documentos recibidas hasta la fecha de la transferencia se siguen haciendo al generador de los documentos. Después de la fecha de una transferencia exitosa de documentos, todas las peticiones de documentos devienen responsabilidad del Archivo.

La infraestructura tecnológica se refiere simplemente al entorno común de hardware y software para ejecutar procesos de conservación. Mi actual infraestructura tecnológica se limita actualmente a un entorno basado en PC, con Windows 2000 como sistema operativo. Este entorno tendrá que actualizarse en un futuro no demasiado distante, porque el volumen de los fondos crece y porque la necesidad de proporcionar acceso al sistema se incrementa.

Una estrategia de conservación se define como una “aproximación coherente y comprensiva a la conservación de un cuerpo de documentos seleccionados para su conservación...” Consiste en objetivos para

- mantener los componentes de los documentos electrónicos y los metadatos y la información relacionados a lo largo del tiempo, expresado en términos de uno o más planes de acción para diferentes clases de componentes digitales, indicando momentos o condiciones para las acciones a ser adoptadas;
- reproducir los documentos en forma auténtica;
- criterios para evaluar la ejecución de la estrategia.

Aquí, la referencia a “componentes” es un reflejo de las conclusiones del Grupo de Trabajo sobre Conservación:

Un documento electrónico se almacena como uno o más componentes digitales. Los componentes digitales no tienen una relación necesaria con los elementos de la forma documental reconocidos en el análisis diplomático de los documentos. Más bien, están determinados tecnológicamente por el modo en que se almacenan los bits y por los métodos (software) que deben aplicarse para reproducir el documento. Reproducir un documento electrónico implica (1) reconstituirlo, esto es, reunir sus componentes digitales si tiene más de uno, o extraer cualquier componente digital almacenado en un fichero físico que contenga más de uno de tales componentes; y (2) presentarlo en forma adecuada².

Un ejemplo familiar es el de una página web que incluya una imagen o fotografía, vista a través de un explorador. La página web normalmente no incluye el fichero o componente de imagen, sino que en lugar de ello incluye texto que indica el modo en que la imagen aparecerá en la página. La imagen misma se almacena como un componente separado de la página web. De modo que hay al menos dos componentes que deben reunirse para reconstituir la página. Nótese que reconstituir la página de este modo también requiere un explorador. Es posible abrir el fichero HTML en una aplicación de procesamiento de texto, y el fichero de imagen en una aplicación de procesamiento de imagen, pero sólo una aplicación con capacidades de explorador será capaz de reconstituir la página del modo en que se pretendió que se presentara. La estrategia de conservación del fichero HTML diferirá en algunos detalles de la del componente imagen complementario.

²InterPARES Preservation Task Force Report, pp. 5-6, disponible en http://www.interpares.org/book/interpares_book_f_part3.pdf, chequeado el 29 de marzo de 2004.

El método objetivo de conservación se refiere al software utilizado para implantar una estrategia de conservación. Esto requiere determinar a qué tipos de documentos o clases de componentes digitales se aplicará el método, y las condiciones de aplicación.

El software utilizado para implantar una estrategia de conservación puede estar íntimamente enlazado a un formato de fichero aceptable para transferencia. Por ejemplo, a los ficheros en formato de valor separado por coma (.csv) puede accederse fácilmente mediante aplicaciones actuales de hoja de cálculo, como Microsoft Excel. En el entorno actual, éste puede ser un modo fiable de reproducir el documento para su acceso. En este caso el método de conservación no requiere ningún tipo de conversión del formato de fichero, aunque la reproducción del documento pueda quedar afectada. Recientemente he comenzado a experimentar con la suite de aplicaciones Open Office³, convirtiendo ficheros formateados en Microsoft Word, Excel, y Powerpoint a formato basado en xml Open Office. Una característica atractiva de Open Office es que proporciona un método para emprender conversiones por lotes. Sin embargo, cualquier decisión de comprometerse en este tipo de método de conservación para todos los documentos necesitaría una comprobación extensa para asegurar que es un método fiable para conservar documentos escritos en francés, o con un formateo inusual, etc.

Finalmente existen los componentes digitales reales, la información complementaria referida a la conservación y reproducción del documento relacionado, y la información sobre la transmisión. Los componentes digitales deben ser transferidos, por supuesto. Es probable que la documentación varíe con los documentos que se transfieren. Es probable que las transferencias de bases de datos requieran diccionarios de datos, que definen los campos de datos e incluso pueden listar el rango de valores para cada campo. Las representaciones gráficas del sistema, incluidos posiblemente los predecesores del sistema, también pueden ser útiles. Para ficheros ofimáticos, las estructuras de directorio, la clasificación de ficheros y las convenciones de denominación de ficheros también pueden ser importantes en términos de documentación complementaria. Finalmente, la información sobre la transmisión, como quién preparó los documentos para su transferencia, el modo en que se prepararon y quién autorizó la transferencia es necesaria para mantener la autenticidad de los documentos.

Estos cinco factores proporcionan una sólida base para desarrollar políticas de transferencia. Estos factores enfatizan las características y necesidades de los documentos en términos de su reproducibilidad y el entorno tecnológico dentro del que debe funcionar el conservador. Quizá de manera más importante, la transferencia de documentos electrónicos debe estar orientada por las políticas de conservación y acceso. La primera, al menos en el contexto de InterPARES, establece explícitamente que el "proceso de conservación digital comienza con el acto inicial de

³Disponible en Sun Microsystems en <http://www.openoffice.org/>, chequeado el 1 de abril de 2004.

almacenamiento y se extiende a lo largo de la reproducción del documento”⁴. El último requisito, i.e., el modo en que se reproducen los documentos (presumiblemente para su acceso) es un factor importante a considerar en el momento de la transferencia. En este momento, el modo en que se acceda a los documentos puede ser el mayor reto para los archiveros, en la medida en que los modelos para acceder a los documentos electrónicos están sólo comenzando a emerger. Es muy posible que los documentos puedan reproducirse en el entorno de laboratorio del conservador, pero no fácilmente para proporcionar acceso a clientes institucionales. Es decir, puede que sea posible reproducir los documentos para que su autenticidad sea evidente a los especialistas, como los archiveros que supervisan la transferencia, pero esto es menos evidente para aquellos que no estén íntimamente familiarizados con las complicaciones del entorno tecnológico y las transferencias.

⁴InterPARES Preservation Task Force Report, p. 5, disponible en http://www.interpares.org/book/interpares_book_f_part3.pdf, chequeado el 29 de marzo de 2004.