

## MODELOS METADATOS PARA LA GESTIÓN DE DOCUMENTOS ELECTRÓNICOS

- Autores:** Eduardo Peis  
epeis@ugr.es
- J.Carlos Fernández-Molina  
jcfernan@ugr.es  
Facultad de Biblioteconomía y Documentación Universidad de Granada
- Resumen:** El imparable incremento de transacciones de todo tipo en la red, las especiales características de los productos de estas actividades y la necesidad de mantener "registro" de muchos de dichos productos está provocando lo que algunos autores han denominado la "reinvención" de los archivos. Las opciones más claras para resolver las necesidades del *recordkeeping* en este contexto pasan por incorporar dichas necesidades al diseño de sistemas empleando para ello un modelo metadatos. Este trabajo analiza algunas aplicaciones y evalúa su adecuación a la nueva situación.
- Abstract:** The unstoppable increment of all sorts of transactions on the Internet, the special characteristics of the products from these activities and the need to keep a "record" of many of these products is bringing about what many authors have called the "reinvention" of archives. Among the most obvious options to resolve the requirements of *recordkeeping* in this context is the inclusion of these requirements in the designing of systems by using a metadata model. The present paper examines some existing applications and evaluates their degree of adjustment to this new situation.
- Palabras clave:** Archivos electrónicos ; Registros electrónicos ; Sistemas de gestión de archivos electrónicos ; Metadatos ; SGML/XML

### Introducción

La "red" existe hoy no sólo como un entramado digital para la difusión informativa, sino que constituye un "espacio" en el que nos comunicamos, enseñamos, jugamos, trabajamos, etc. Sobre este "espacio" se hacen negocios de gran volumen, los gobiernos reciben y proporcionan información a los ciudadanos, los centros de formación imparten sus enseñanzas y se relacionan con los estudiantes, etc.

Todas estas actividades, que en su gran mayoría no siguen los cánones burocráticos tradicionales, producen "documentos" que pueden ser reflejo de un instante, que pueden ser un conjunto de muchos documentos multimedia (¿puede ser un documento la "navegación" desarrollada en un momento determinado, buscando algo concreto?), que son modificables, que no tienen consistencia física (¿es el documento lo que vemos en pantalla o "salvamos" en el disco?). Por otra parte, de muchas de estas actividades es necesario

dejar constancia, mantener testimonio de su ocurrencia: ¿es necesario hacer "registro" de la comunicación electrónica de la resolución de una convocatoria de plazas de profesorado?, ¿es importante conservar testimonio de unas actas formalizadas en la web?

Los registros electrónicos deben ser tratados, deben ser gestionados para asegurar la evidencia de las actividades de las que proceden.

El sistema informático de cualquier organismo, en sus capas administrativas, se puede considerar como un sistema de gestión electrónica de documentos. Estos sistemas son cada vez más efectivos y posibilitan una gestión eficaz, pero tan sólo de aquellos documentos que han sido creados como tales. Algunas de sus características funcionales (registro, acceso o recuperación) podrían coincidir con las necesidades de un sistema de gestión de archivos, pero éste debe permitir, además, como mínimo: determinar un documento como documento de archivo, administrar el cuadro de clasificación, controlar los periodos de retención y disposición...

La tendencia tecnológica actual, por lo menos en las empresas, es integrar las funcionalidades de un sistema de gestión de archivos en el sistema de gestión electrónica de documentos, bien utilizando aplicaciones diferentes o bien añadiendo funciones propias de la gestión de documentos de archivo (mediante módulos o adds-on) a sus sistemas.

Paradójicamente, la esencia de esta integración se basa en la posibilidad de que no sea necesaria la intervención de un profesional de archivo (sería el propio usuario, por ejemplo, el que crearía el "perfil" del documento y el que determinaría su declaración como registro), pero se construye sobre principios archivísticos tradicionales (cuadros de clasificación, calendarios de disposición, etc.).

Los sistemas de gestión de archivo actuales posibilitan una automatización muy rentable del sistema de archivo, de hecho algunos de ellos han sido diseñados por profesionales de los archivos, pero la gestión de aquellos registros electrónicos que desde un principio no han sido concebidos como tales puede resultar impracticable.

Lo cierto es que, tanto con los sistemas de gestión de archivo "tradicionales" como con los sistemas de gestión electrónica de documentos adaptados, de muchas de las actividades a las que aludíamos al comienzo de este trabajo es muy difícil mantener registro y, por otra parte, es muy complicado que estos sistemas integren los principios y prácticas archivísticas tradicionales y al mismo tiempo se enfrenten a estos "documentos". Y es que el impacto del ciberespacio sobre los sistemas de gestión de archivos es y será enorme.

## **El impacto tecnológico**

Los principales cambios que pueden afectar a los sistemas de gestión/tratamiento están relacionados con el cambio en la forma del documento y la modificación de los métodos de trabajo. Estas modificaciones dificultan la aplicación de los principios y prácticas tradicionales a este tipo de documentos.

Para Schamber (1) un documento electrónico tiene una serie de características que lo diferencian del tradicional: es fácilmente manipulable, enlazable interna y externamente, rápidamente transformable, intrínsecamente localizable, instantáneamente transportable e infinitamente replicable.

De hecho, el documento electrónico consta de una serie de señales digitales y, por lo tanto, tiene pocos o ninguno de los atributos físicos del documento tradicional. Los atributos físicos del documento electrónico, que incluirían la forma o el tipo de material cuando es visualizado en pantalla o impreso, son en gran medida una función del software y están separados del contenido informativo o del contexto del documento. Las relaciones lógicas de un documento dependen de los atributos físicos y de otra información contextual que genera el software (2).

Respecto al contexto productivo, la tecnología de la información proporciona al trabajador que opera en una estructura de tipo jerárquico la posibilidad de comunicarse horizontalmente, fuera de los canales "normales", y de desarrollar una actividad o cooperar sin la intervención de una autoridad jerárquicamente superior (3). Los actuales procedimientos archivísticos no se prestan fácilmente a la transmisión horizontal de documentos electrónicos (2). La descentralización promueve la colaboración y la difusión de la información sin límites de tiempo o espacio. Es lo que se ha denominado "espacio virtual de trabajo común" (3). Es más, estas modificaciones implican la transformación, tanto de las relaciones entre el proveedor y el usuario del servicio, como de los recursos de información que están en el corazón de esta asociación (4). Así, los ficheros de datos dejan de ser estáticos en sí mismos: algunos son creados "a demanda", basándose en las necesidades de los usuarios o los proveedores de servicios.

Durante las dos últimas décadas, los archiveros se han enfrentado a los cambios que representa la tecnología electrónica. Los primeros autores que abordaron el tema, ciñendo el análisis al producto, creyeron que los archiveros podrían aplicar la teoría y la práctica archivística tradicionales a los documentos en formato electrónico. En los últimos años, sin embargo, algunos investigadores sostienen que la propia naturaleza de los registros electrónicos y el contexto en que se producen requieren de los archiveros la adopción de nuevas ideas que podrían cambiar, e incluso eliminar, los principios archivísticos tradicionales.

## La respuesta

Para hacer frente al impacto de la tecnología de la información, se ha desarrollado un movimiento teórico-práctico que busca respuestas aplicables a la gestión de los documentos electrónicos. Se propone la sustitución del tradicional "ciclo de vida de los registros" por el de "flujo continuo de registros" (5). El marco de trabajo sobre el que se asienta este movimiento es el modelo conocido como *records continuum*.

Este modelo está construido sobre la afirmación de que la gestión del documento es un proceso continuo desde el momento de creación. Los conceptos relativos a dicha gestión pertenecen a cuatro dimensiones que no son relativas a la edad de los registros, sino que tienen que ver con el punto de vista del observador (6).

En este marco conceptual y abordando la complejidad de los registros electrónicos, considerándolos como algo nuevo, importantes autores como Cook (7), Cox (8), Dollar (2) o Hedstrom (9), han comenzado a formular nuevas ideas para tratarlos. Influidos por las ideas de David Bearman (6, 10, 11), estos autores argumentan, en términos generales, que los archiveros deberían modificar su centro de atención, desde el contenido de un registro a su contexto; desde el registro en sí mismo a la función de dicho registro; desde la preservación y acceso hacia la intervención en el proceso de creación. Se trataría de una nueva tendencia, una actitud estratégica, que algunos han denominado (de forma poco rigurosa) el "nuevo paradigma" (12).

Aunque el consenso profesional con respecto a estas cuestiones es escaso todavía, lo que sí parece claro es la inadecuación de considerar las cuestiones archivísticas al final del ciclo de vida de la documentación, para demostrar la urgencia de forzar la reestructuración de los principales temas de la gestión de documentos y archivos. Así, para Hedstrom y Wallace (13), "la mayoría de los archiveros están ahora de acuerdo en que las cuestiones del acceso y la preservación futuras deben ser parte integral del diseño de las políticas y sistemas de información. En teoría, la gestión de documentos y archivos, que tradicionalmente ha sido tratada a posteriori, debe ser catapultada hacia las consideraciones previas de diseño de políticas, sistemas y aplicaciones."

### **Aplicaciones. Modelos metadatos para la gestión de documentos electrónicos.**

En un contexto tecnológico en el que resulta prácticamente imposible desarrollar las funciones archivísticas tradicionales, los archiveros deberían, por ejemplo, determinar las maneras de influir en las normas tecnológicas que apoyan los sistemas de información. Esta influencia debería ejercitarse en la fase de diseño de sistemas, con el fin de conseguir regular legislativamente las operaciones que tienen relación con la creación, el uso, la difusión y preservación de la información (13). Las normas ofrecen a los archiveros un medio para

estipular los criterios archivísticos a respetar en dichos diseños para los documentos electrónicos (criterios que serían, lo que Bearman [6] denomina “requerimientos”).

Por otra parte, es necesario encontrar la manera en que, mediante una valoración funcional apriorística, los registros puedan ser autodescritos y autogestionados por el propio sistema. Las aplicaciones desarrolladas hasta el momento, con mayor o menor éxito en su implementación, utilizan para ello modelos metadatos, regulados por las normas rigurosas aludidas en el párrafo anterior, que pueden ser integrados en el sistema.

En mayo de 1996 la *Information Management Standards and Practices Division* de los Archivos Nacionales de Canadá diseñó una estrategia denominada *Electronic Work Environment* (EWE). El modelo EWE requiere de las agencias la creación de registros que documenten adecuadamente sus actividades y la seguridad de que estos registros permanecerán disponibles, comprensibles y utilizables mientras sean necesarios. Su principio básico es que “la forma más efectiva de gestionar registros en el EWE es incorporar las necesidades del *recordkeeping* en los diseños de los procesos comerciales automatizados” (14).

Para determinar el potencial de este modelo se desarrolló un prototipo. Su interfaz ha sido diseñado con una serie de iconos que representan las actividades empresariales. Los registros son esencialmente clasificados y sustancialmente autodescritos mediante la conexión a metadatos en el momento de la creación, incluyendo detalles sobre sus funciones, actividad, uso, antecedentes, estatus, entorno software, organismo, permisos, retención y género o tipo (14).

También en 1996 vio la luz la primera norma nacional para la gestión de registros electrónicos en Australia (AS-4390) (15). La norma recomienda una metodología para el diseño o rediseño de un sistema para la gestión de registros que incluya esquemas para el control de registros (registro, clasificación, indización y flujos de uso), decisiones de valoración para aquellos registros a capturar por el sistema, determinación de su valor a largo plazo para obtener programas de disposición y necesidades de almacenamiento para uso operativo y mantenimiento a largo plazo. Se han realizado esfuerzos para desarrollar una norma internacional en este sentido por parte de la ISO. Los *Australian National Archives* han adoptado explícitamente la metodología recomendada por la AS-4390 en su política para la gestión de registros electrónicos (16).

El resultado ha sido el desarrollo de una metodología, recogida en el manual *Designing and Implementing Recordkeeping Systems* (DIRKS) (17), que permite a las agencias productoras contribuir al diseño e implementación de sistemas mediante la concreción de las necesidades funcionales a incorporar a tales sistemas a través de un conjunto metadatos regulados por la

*Recordkeeping Metadata Standard for Commonwealth Agencies* (versión 1.0) (18).

La primera parte de la norma explica el propósito y la importancia de aplicación del conjunto de metadatos normalizado y detalla el alcance y las características de la norma. Dichas características incluyen: flexibilidad de aplicación; repetibilidad de los elementos de datos; extensibilidad para permitir la gestión de necesidades específicas; interoperatividad entre sistemas; compatibilidad con normas metadatos relacionadas, incluyendo la norma *Australian Government Locator Service* (AGLS), derivada de *Dublin Core* (19); e interdependencia de los metadatos a nivel de sub-elementos.

La segunda parte proporciona una detallada descripción de los 20 elementos de datos (agente, gestión de derechos, título, materia, descripción, lengua, relación, cobertura, función, fecha, tipo, nivel de agrupación, formato, identificador de registro, historia de gestión, historia de uso, historia de conservación, localización, disposición y autorización) y 65 sub-elementos, definiéndolos en relación a su propósito y fundamentación. Para cada elemento y sub-elemento la norma proporciona indicación de su ámbito de aplicación, obligatoriedad, condiciones de uso, valores asignados y estructura aprobada.

Los elementos pueden ser divididos en seis categorías o niveles: registro, términos y condiciones, estructura, contexto, contenido e historia de uso, que reflejan su función en la autenticación y la gestión a largo plazo de los registros y que coinciden con el modelo de referencia metadata desarrollado por la Universidad de Pittsburgh como parte de sus *Functional Requirements for Evidence in Recordkeeping*.

Muchas de las aplicaciones (o proyectos de aplicaciones) desarrolladas hasta el momento podrían enmarcarse en una estrategia global que estaría muy relacionada con lo que Bearman (10) denomina “imponer requerimientos funcionales para el *recordkeeping*”. Efectivamente, en el marco del *Electronic Records Project* de la Universidad de Pittsburgh, dirigido por Bearman, y concluido en 1996 (20), fueron definidos los ya citados “*Functional Requirements for Evidence in Recordkeeping*”, que pueden ser traducidos a especificaciones formales para sistemas de gestión de documentos. Estas especificaciones son satisfechas por metadatos concretos introducidos en y mantenidos por los sistemas, y son asignados -o conectados- a los registros (21).

La especificación de estos requerimientos funcionales, expresada en “reglas de producción” o definiciones lógicas de atributos simples observables, concreta veinte propiedades que han sido identificadas en la sociedad universitaria como propiedades fundamentales de los registros.

Las características necesarias y suficientes para los datos que pretenden ser registros se concretan en un conjunto de metadatos que, cuando están presentes, satisfacen la especificación. Si estos metadatos están

insolublemente conectados a -y retenidos con- los datos asociados con cada transacción administrativa, estará garantizado que el objeto de datos será utilizable con el tiempo, accesible sólo bajo los términos y condiciones establecidas por sus creadores, y tendrá las propiedades requeridas para ser considerado como evidencia (6).

Para implementar los requerimientos funcionales es necesaria una norma técnica rigurosa. En este proyecto se desarrolló dicha norma y se denominó "*Reference Model for Business Acceptable Communications*" (abreviada como BAC). Este modelo de referencia intenta regular los requerimientos a través de una estructura general de objetos metadata encapsulados (*metadata encapsulated objects* –MEO-), que corresponderían al modelo de metadatos incrustados considerado por Weibel (22), agrupando estas categorías de metadatos y elementos para conseguir modularidad funcional y organizándolos en los seis niveles *cluster* anteriormente citados, que permiten desarrollar el proceso técnico (20).

Para salvar los registros de un sistema propietario y para capturar todos los metadatos relevantes, Bearman (6) propone utilizar una norma de aceptación general como SGML (*Standard Generalized Markup Language*) y el desarrollo de una DTD (*Document Type Definition*) que refleje la estructura general de objetos metadata encapsulados que presenta el *Reference Model for Business Acceptable Communications*. Basar la estructuración lógica de este modelo en una DTD SGML, es una elección adecuada, ya que la norma puede ser referenciada en los metadatos, de manera que el contenido de los datos puede ser accesible conociendo la norma. Esta opción presenta una doble ventaja: consistencia de las definiciones y facilidad para la "migrabilidad".

## Conclusiones

La tecnología de la información está produciendo un impacto "revolucionario" en el mundo de los archivos y la gestión de documentos. El impacto tecnológico, que incide sobre todo en la modificación de la forma del documento y de los métodos de trabajo, obliga a diseñar una nueva estrategia que puede modificar (e incluso eliminar) algunos de (o todos) los métodos tradicionales de tratamiento para hacer frente a los documentos electrónicos

Conceptualmente hablando, una estrategia posible consiste en modificar el centro de atención desde el contenido de un registro a su contexto; desde el registro en sí mismo a la función de dicho registro; y desde la preservación y acceso hacia la intervención en el proceso de creación.

En la práctica, para cumplir con esta estrategia es necesario construir modelos completos de gestión/archivo de registros electrónicos, e incorporar las necesidades funcionales de dichos modelos en el diseño de nuevos procedimientos laborales y en los sistemas tecnológicos que los soporten. Y esto ha de llevarse a cabo en la fase de diseño de sistemas.

En todos estos modelos subyace como estructura lógica para su desarrollo un modelo metadatos, que permite incorporar al sistema las necesidades funcionales eficaces para gestionar registros electrónicos, traduciéndolas a especificaciones lógicas reguladas por un marco normativo riguroso. Bearman, en su momento, propuso emplear SGML como metalenguaje adecuado para representar la estructura del modelo metadata a emplear. Sin embargo, el avance continuo de las posibilidades tecnológicas nos permite afirmar que la herramienta tecnológica ideal para ello, hoy día, es XML (23) (y sus especificaciones adicionales, XLink, XSL), que presenta muchas de las posibilidades de SGML siendo viable su uso, además, utilizando aplicaciones software no específicas. En nuestra opinión, una arquitectura ideal para construir el modelo (que además emplea XML) sería RDF "Resource Description Framework".

## Referencias

1. SCHAMBER, L. What is a document? Rethinking the concept in uneasy times. *Journal of the American Society for Information Science*, 1996, vol. 47, nº 9, p. 669-671
2. DOLLAR, C. M. *Archivistica e Informatica: l'impatto delle tecnologie dell'informazione sui principi e sui metodi dell'Archivistica*. Macerata: Pubblicazioni dell'Università, 1992
3. ZUBOFF, S. *In the age of the smart machine: the future of work and power*. New York: Basic Books, 1988
4. KESNER, R. M. Information resource management in the electronic workplace: a personal perspective on "Archives in the Information Society". *American Archivist*, 1998, vol. 61, nº 2, p. 70-87
5. UPWARD, F. The continuum: principles, structures and dualities. *Archives and Manuscript*, 1997, vol. 25, nº 1, p. 27-46
6. BEARMAN, D. A. Item level control and electronic recordkeeping. *Archives & Museum Informatics*, 1996, vol. 10, nº 3, p. 214-217
7. COOK, T. What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria*, 1997, vol. 43, nº 2, p. 17-63
8. COX, R. J. The record in the information age: a progress report on research. *The Records & Retrieval Report*, 1996, vol. 12, nº 1, p. 1-16
9. HEDSTROM, M. Electronic archives integrity and access in the network environment. *American Archivist*, 1995, vol. 58, nº 3, p. 312-324

10. BEARMAN, D. A. Archival strategies. *American Archivist*, 1995, vol. 58, nº 4, p. 380-413
11. BEARMAN, D. A.; HEDSTROM, M. Reinventing archives for electronics records: alternative service delivery options. En: M. Hedstrom, ed. *Electronic records management program strategies*. Pittsburgh, PA: Archives & Museum Informatics, 1993, p. 123-156
12. COOK, T. Electronics records, paper minds: the revolution in information management and archives in the post-custodial and post-modernist era. *Archives and Manuscript*, 1994, vol. 22, p. 300-328
13. HEDSTROM, M.; WALLACE, D. And the last shall be first: redordkeeping policies and the NII. *Journal of the American Society for Information Science*, 1999, vol. 50, nº 4, p. 331-339
14. McDONALD, J. Research issues in interfaces for the capture of business processes. *Electronic Records Meeting (Pittsburgh, May, 29, 1997)*. Pittsburgh: SAA, 1997
15. Australian Standard AS 4390–1996, *Records Management*. Canberra: AS, 1996
16. ROBERTS, D. *The new Australian records management standard*. Paper presented at the annual meeting of the National Association of Government Archives and Records Administrators (NAGARA), Sacramento, CA (1997, July 19). Sacramento: NAGARA, 1997  
<http://www.records.nsw.gov.au/rk/sacramento.htm> (febrero, 2001)
17. NATIONAL ARCHIVES OF AUSTRALIA. *Designing and implementing recordkeeping systems*. Canberra: NAA, 1999  
<http://www.naa.gov.au/recordkeeping/dirks/summary.html> (febrero, 2001)
18. NATIONAL ARCHIVES OF AUSTRALIA. *Recordkeeping Metadata Standard for Commonwealth Agencies, Version 1.0*. Canberra: NAA, 1999  
<http://www.naa.gov.au/recordkeeping/control/rkms/summary.htm> (febrero, 2001)
19. Dublin Core Metadata. *Dublin Core Web Site*.  
<http://purl.oclc.org/DC/about/workshop.htm> (febrero, 2001)
20. Pittsburgh Electronic Records Project Homepage.  
<http://www.lis.pitt.edu/~nhprc> (febrero, 2001)
21. BANTIN, P. C. Developing a strategy for managing electronic records. The findings of the Indiana University Electronic Records Project. *American Archivist*, 1998, vol. 61, nº 4, p. 328-364

22. WEIBEL, S. L. (1997). The Dublin Core: a simple content description model for electronic resources. *Bulletin of the American Society for Information Science*, 1997, vol. 24, nº 1, p. 9-11
  
23. W3C-XML (1998). *World Wide Web Consortium. Extensible Markup Language (XML) 1.0 Recomendación 10 de febrero de 1998.*  
<http://www.hispalinux.es/http://www.w3.org/TR/1998/REC-xml-19980210.html> (febrero, 2001)