

## La Gestión de Contenidos de Empresa y Código Abierto

La gestión de contenidos de empresa (ECM) es constantemente una de las diez iniciativas más importantes de funcionarios de información principales de Fortune 1000. Durante la década pasada, la ECM ha cambiado de algo adicional a ser una necesidad. Esto es porque los contenidos en sus muchas diversas formas desempeñan un papel imprescindible en el funcionamiento del negocio. Los contenidos se utilizan para presentar productos y para llevar a cabo el comercio con los clientes. Los documentos de contenidos prueban que los procesos del negocio funcionen bien y estén mejorando constantemente especialmente en ambientes regulados y con requisitos de la conformidad como Sarbanes-Oxley. Los contenidos son el medio en el cual los diseñadores, científicos, el marketing y la gerencia planean y tratan nuevos y existentes mercados. Los contenidos informan a cada uno que esté involucrado en la cadena de valor de qué está sucediendo y de cómo se puede mejorar. Las leyes de Moore y de Gilder se han asegurado de que el flujo de la información y de los contenidos crecen a medida que las computadoras desktop puedan almacenar, producir y descargar más. Los contenidos ya cubren páginas Web, documentos de la oficina, imágenes escaneados, correos electrónicos y archivos electrónicos.

El término gestión de contenidos se originó alrededor de 1996 y se le acreditaría a compañías como Vignette e Interwoven por haberlo popularizado. Sin embargo, en 1996 David Yockelson de Meta Group planteó la pregunta retórica de ¿qué es la diferencia entre la gerencia de los documentos, la gerencia de la imagen y la gerencia de contenidos cuando todo el contenido será accesible de un navegador del Internet. Durante los siguientes años, estas áreas dispares de la gerencia de información no estructurada convergieron y crearon el mercado de la Gestión de Contenidos de Empresa, un término que había existido desde alrededor de 1998, pero realmente solo había ganado algún significado en el 2001 y ha ganado enorme impetu durante los dos años más recientes.

La gestión de contenidos verdaderamente ahorra dinero y aumenta productividad. Al automatizar los procesos de tocar los contenidos y al simplificar el acceso de la información del desbordamiento de la empresa y del Internet, la gestión de contenidos reduce el tiempo que el trabajador emplea en manipular el papel y hacer su trabajo. Muchos de los procesos de la ingeniería y de la calidad en empresas no podrían existir sin la gestión de contenidos debido al puro volumen de contenidos que se maneja. El proceso de manejar páginas Web corporativas, el intranet y el extranet ahora no se pueden hacer sin la gestión de contenidos, como han llegado a ser el conducto principal de la comunicación, reponiendo el papel y el correo electrónico.

## Aplicaciones del ECM

Como uno de los fundadores de Documentum, he visto muchos tipos diferentes de programas de gestión de contenidos durante los últimos 15 años, particularmente en ambientes empresariales. La mayoría se centran en solucionar problemas importantes con los volúmenes grandes de los contenidos que se manejan manualmente. Algunos usos típicos incluyen:

- \* Programas que se utilizan para cumplir con reglas, políticas y regulaciones. Programas que se aseguran de que los procesos estén documentados, los materiales están marcados y versionados claramente, y se revisa el uso.
- \* Los portales del conocimiento hacen más eficiente el proceso de poner lo que los expertos en una empresa han escrito en las manos de la gente que necesita esa información a pesar de donde se encuentre. Los portales del conocimiento proporcionan interfaces de la búsqueda y el hojear en los contenidos relevantes para las actividades como inteligencia competitiva, ventas y diseño.

- \* La gestión colaborativa de contenidos tiende ser orientada hacia proyectos, reuniendo a un equipo de gente centrada en un objetivo común. Los contenidos utilizan una metáfora de un cuarto, el espacio o el sitio se usa para indicar que aquí es donde el equipo se reúne y mejora los contenidos con la comunicación, discusión y resolución de problemas.
- \* La gestión de documentos es el manejar los contenidos que se usan en las operaciones de negocio cotidianas, particularmente documentos de la oficina. Los documentos típicos que se manejan son documentos del tratamiento de texto, hojas de cálculo, presentación y documentos escaneados.
- \* La gestión de los contenidos del Web es el término conectado más con la gestión de contenidos, pero es una porción pequeña de los requisitos de contenidos de una empresa. La gestión de los contenidos del Web ha estado en gran parte convertido en producto básico por alternativas de código abierto. Sin embargo, los requisitos especiales de la empresa se vuelven a usar a través de muchos sitios, ayuda para los Internet corporativos, extranet de negocio a negocio, espacios Web multilingües, espacios Web de los medios publicitarios, y gestión de derechos digitales [digital rights management], y activos más eficazmente.

El crecimiento y la creciente complejidad de estas aplicaciones ha estimulado la industria de la gestión de los contenidos para ser una de las pocas clases del software de la empresa haber crecido realmente durante la recesión y durante los últimos cinco años. Esto también ha conducido a compañías principales de la industria, EMC Documentum, IBM, Microsoft Sharepoint, Filenet, OpenText, Vignette y Interwoven a adquirir componentes que necesitaron para tratar las aplicaciones anteriores. Cualquier cosa que fuera complementaria a la gestión de los contenidos de empresa se convirtió en parte de ECM con estas adquisiciones. Las compañías principales todavía están ampliando la lista, mientras que compañías menos principales todavía están tratando de alcanzarlas.

## ¿Qué es La Gestión de Contenidos de Empresa?

El sistema de los servicios que constituyen la Gestión de Contenidos de Empresa es grande y comprensivo. [AIIM (Association for Information and Image Management)] está intentando catalogar e identificar los servicios más críticos que constituyen el ECM como parte de su comité de IECM (Interoperable ECM). (<http://www.aiim.org/estándares.¿cAsp?ID=29284>) Este trabajo ya está comenzando, pero he juntado mi lista de los servicios que creo que proporcionan la coherencia y la interoperabilidad de componentes y de los contenidos en el contexto de la gestión de contenidos en un nivel de la empresa. Estos servicios incluyen:

- \* Servicios de la Biblioteca, cuál es un término que abarca las operaciones básicas de los contenidos que distinguen a la gestión de los contenidos de la gestión ordinaria de la base de datos. Los servicios de la biblioteca incluyen el acceso básico, almacenar, ingresar, [check-in, check-out], cerrar, control de la versión, manejar metadatos, y el organizar los contenidos en estructuras de carpeta. El auditar de quién tiene acceso y actualiza la información se pueden considerar una parte de servicios de la biblioteca.
- \* Los Servicios de la Organización proporcionan los esquemas de la clasificación y los taxonomías que permiten que la información sea hojeada más fácilmente en categorías múltiples y a confinar búsquedas. La forma más básica de organización es el uso de carpetas para clasificar la información.
- \* La búsqueda y la recuperación abarca texto completo, metadatos y búsquedas de la categoría. El texto completo ha sido definido en gran parte por los vendedores de la

búsqueda del Internet como Google, cuáles han fijado la expectativa de los usuarios en facilidad de empleo y velocidad.

- \* Los servicios de la colaboración son relativamente nuevos y conducidos por la convergencia del conjunto de programas en el mismo espacio que la gestión de los contenidos. La colaboración generalmente se enfoca en equipos y gira alrededor del comentar y anotar de los contenidos, seguir el desarrollo de proyectos, distribución de recursos, y el seguir el desarrollo de asuntos.
- \* Los servicios del flujo del trabajo implementan procesos del negocio y aseguran la ejecución sin complicaciones al proveer el equilibrio del volumen de trabajo y seguir el desarrollo de cuáles tareas se han realizado cuáles son atrasadas. Actualmente, los servicios del flujo del trabajo están experimentando cambios importantes como resultado de la industria que estandariza en el estándar de BPEL4WS.
- \* Los servicios del ciclo de vida manejan las políticas a largo plazo que tratan y son cruciales para decretar políticas del archivar y eliminar de la gestión de archivos. Son también necesarios para el control de la visibilidad de los contenidos en proceso que generalmente la industria llama efectividad.
- \* Los servicios de seguridad controlan permisos de contenidos que incluyen el simple leer, escribir y cancelar hasta el copiar, imprimir, acceso de las características especiales y delegación de la propiedad. Los servicios de seguridad modernos proveen listas de control basadas en papeles con la habilidad no sólo designar quién tenga acceso ver los contenidos pero también excluir otros de tomar medidas por causa de los contenidos.
- \* Servicios de Transformar / Publicar convierte los contenidos y los componentes de los contenidos en otros formatos y junta componentes para formar documentos completos y páginas Web. La transformación es útil para crear contenidos listos para el Web de otros formatos.
- \* Servicios de Acontecimientos/ distribución son para proporcionar cambios y actualizaciones de los contenidos en un depósito y para permitir que los usuarios suscriban a esos cambios. [RSS] se está convirtiendo en un mecanismo popular para proporcionar esos cambios.
- \* Los servicios de la Alta-Disponibilidad entregan a la gestión de contenidos la habilidad de proporcionar mayor aseguramiento del acceso y la actualización de los contenidos y se proporcionan a través de: reserva y restauración; aprovisionamiento de un [hot standby]; o réplica a otro sitio o sistema para acceso geográfico más cercano o recuperación del desastre.
- \* Los servicios de la administración proporcionan una vista como la de un tablero de instrumentos y un panel de control para el sistema de gestión de contenidos activo; supervisa las actividades de servicios de contenidos; y se asegura de que se conformen con los niveles de disponibilidad que requiere la empresa.

## **El Uso de la Gestión de Contenidos de Empresa**

La manera de que los usuarios obtienen acceso a los servicios de ECM dependen de la aplicación que se ejecuta. La mayoría de los usuarios no quieren saber los detalles de estos servicios y preferirían que permanecen fuera de vista y automáticos. Hay cuatro paradigmas principales con que los usuarios están cómodos y quisieran que la gestión de contenidos de empresa se conformara con estos paradigmas. La gestión de contenidos de empresa ha intentado adaptarse a estos paradigmas y ha tenido éxito mezclado. Estos paradigmas incluyen el hojear archivos, el buscar de estilo de Internet, portales y correo electrónico.

- \* El navegador del archivo y el explorador con su compañero las unidades compartidas son el paradigma más viejo y sencillo con el cual los usuarios sienten cómodos y es el lugar más fácil a almacenar, entrar y compartir contenidos de todos tipos. Las unidades compartidas están disponibles todo el tiempo, aun cuando alguien viaja. Muchos vendedores del ECM proveen un cliente del estilo de Windows Explorer que tiene características de la gestión de contenidos, pero generalmente a los usuarios no les gusta estos clientes porque los requiere utilizar algo diferente de su ambiente estándar de su computador personal. Los protocolos de WebDAV y de DeltaV proporcionan algo de la funcionalidad de [versioning] y de fijación del ECM, pero no al acceso fuera de línea.
- \* Google y la búsqueda del Internet es un paradigma que ha tenido un impacto enorme en la gestión de contenidos de creando una expectativa en usuarios de empleo fácil y tiempo de reacción rápido. Los usuarios quieren poder hacer búsquedas en sus almacenes por lo menos tan rápidamente como en el Internet. Por lo tanto, muchos vendedores del ECM han incorporado buscadores que imitan la capacidad de Google y han aumentado esta experiencia de la búsqueda con el auditoría, control de acceso, el cambiar las historias, el procesar la información y la capacidad de invocar flujos del trabajo.
- \* Los portales de la empresa consolidan la selección heterogénea de aplicaciones y mucha de la información que los usuarios finales desean son contenidos de los almacenes de contenidos. Los portales simplifican el uso y la gerencia de las operaciones de la gestión de contenidos y son el interfaz primario para usar algunos de los servicios más complejos tales como el flujo de trabajo, colaboración, ciclo de vida, seguridad y administración.
- \* El correo electrónico es a menudo la herramienta que escoge el trabajador de conocimiento y la mayoría de los procesos del negocio basados en conocimiento de hecho se funcionan a través del correo electrónico. La mayoría de los usuarios preferirían dirigir el flujo del trabajo con los mensajes del correo electrónico que contienen instrucciones, enlaces y herramientas para realizar la tarea que se requiere del usuario.
- \* [RSS] es un paradigma nuevo del consumo de información que está emergiendo. El protocolo [Really Simple Syndication] proporciona rótulos de [XML] de la información que ha cambiado y muchas herramientas, incluyendo navegadores del web y los clientes del correo electrónico, pueden procesar corrientes de [RSS].

## Desafíos que Hacen Frente al Mercado de ECM

Una encuesta conducida en 2003 por el [Information Architecture Institute (Instituto de la Arquitectura de la Información)] ([http://iainstitute.org/pg/the\\_problems\\_with cms.php](http://iainstitute.org/pg/the_problems_with cms.php)) encuestó usuarios actuales del software de la gestión de contenidos para identificar los desafíos que sus organizaciones tenían con la gestión de contenidos. El artículo número uno que los respondedores identificaron era el coste de software comercial. Esto fue seguida por el proceso pobre de la migración de contenidos existentes y la inhabilidad de adaptar el software a procesos existentes y diseños y la complejidad de modificar el sistema para requisitos particulares. Mucho del resto del examen hizo hincapié en la complejidad y la dificultad de usar sistemas existentes. Aparte de coste, los sistemas actuales del ECM simplemente son demasiado difíciles de usar.

El profesor Clayton Christensen de Harvard en su libro, *El Dilema Del Innovador*, observó que las industrias maduras llegan a un punto en el cual la funcionalidad basta y los clientes desean la conveniencia, confiabilidad y precio bajo. La encuesta por el [IAI] y las opiniones

de los analistas de la industria indican que el ECM está llegando a este punto. El mercado de la gestión de contenidos de empresa ya tiene \$2.5 billones de dólares de ingresos con más que mitad de esos ingresos de servicios. La funcionalidad principal ya se está convirtiendo en producto básico y los vendedores están consolidando con la mayoría de los vendedores proporcionando más o menos la misma funcionalidad, la cual es demasiado para la mayoría de los propósitos. Los clientes están vacilando comprar características adicionales, y los vendedores del ECM compensan con los servicios para aumentar los ingresos. Los usuarios finales aun están rechazando utilizar a la gestión de contenidos, prefiriendo almacenar los contenidos en un sistema de ficheros compartido debido a su facilidad de empleo.

La industria del ECM también ha sido plagada por una escasez de aplicaciones empaquetadas y de programadores capaces de incorporar aplicaciones comparadas al mercado de la base de datos. La razón primaria de esto ha sido una falta de la estandarización y de la interoperabilidad. Cada vendedor del ECM tiene su propio API (interfaz de programación de aplicaciones), herramientas e interfaces que lo hacen difícil para un programador saber más que un par de sistemas diferentes e imposible escribir una aplicación para más de uno. Las primeras medidas hacia la estandarización se han tomado con WebDAV y JSR-170, un interfaz de Java a los sistemas del CMS, pero éstos serán limitados a Java y al desarrollo técnico. Esta misma carencia de la interoperabilidad ha incitado a muchos funcionarios de información principales pedir un enfoque de la arquitectura que se orienta al servicio para la gestión de contenidos y una necesidad para los servicios estándares del web permitir ambientes de Java, .NET y LAMP tener interoperabilidad.

A pesar de crecimiento positivo de 15% por año, la industria de la gestión de contenidos de la empresa ni se acerca al tamaño de la industria de la base de datos. Los obstáculos que previenen que el ECM crezca mucho más son costo más bajo, facilidad de empleo, conveniencia y el crecimiento de la población de desarrolladores para desarrollar aplicaciones contenidos. El potencial existe para hacer que el mercado del ECM sea tan grande como el mercado de la base de datos, pero no bajo el modelo actual del desarrollo de mercado. Una industria que cuenta con añadir más funcionalidad para subir los precios y generar mas ingresos con el servicio no puede proporcionar el modelo que es más simple, más barato, y más conveniente.

## **El Código Abierto y la Gestión de Contenidos de Empresa**

El código abierto ha emergido como un nuevo modelo para desafiar el modelo tradicional del software de la empresa. Las mismas fuerzas que han conducido la extensión de la gestión de contenidos de la empresa, el Internet y aumentado la capacidad de computadoras, han conspirado deshacer el modelo del crecimiento del software de la empresa. Los clientes ya no requieren una fuerza de ventas para descubrir o ser convencidos de que necesitan cualquier software particular. Los clientes no necesitan una cadena de oferta compleja proveerles software cuando pueden bajar soluciones completas ellos mismos de su escritorio. Más importantemente, la conexión del cliente al desarrollador no requiere a montones de gente anotar ideas, deseos y quejas y traducirlos para que los desarrolladores del software los conviertan en soluciones de software y nuevas versiones.

El código abierto como un modelo de la distribución y un modelo de la innovación es la evolución natural de la industria del software de la empresa. Establece un mecanismo muy eficaz de ventas, marketing y de la distribución que permite que los clientes descubran el software que necesitan para sí mismos. Pueden acceder y probar el software antes de invertir en el. La industria de la gestión de contenidos de empresa es particularmente vulnerable a esta tendencia puesto que los clientes se están desencantando con el precio de ECM. La marca que los participantes han desarrollado a lo largo de los años les protege, pero los participantes más pequeños deben adaptarse o desaparecer. El código abierto se ha llegado a ser el medio nuevo de la producción e innovación que presiona a todos con la excepción de los fabricantes más grandes.

El código abierto ha producido docenas de sistemas de gestión de contenidos durante los últimos varios años. Casi todos estos sistemas son sistemas de la gestión de contenidos del Web, una faceta de la gestión de contenidos de empresa que se convirtió en producto básico hace años. Algunos de los sistemas que han desarrollado uso considerable en los últimos dos años son algunos de los sistemas basados en PHP: Mambo, XOOPS, Nukes y Drupal. Estos sistemas son un sistema del código de PHP que alguien cargaría en un proveedor de servicios Internet y el [CMS (Sistema Monitor Conversacional?)] se convierte en una página de Internet. Son sencillos y tienen comunidades activas y vibrantes que contribuyen todas clases de extensiones para cosas como foros, [wikis], y calendarios. Sin embargo, su funcionalidad se limita generalmente al contenido del Web y casi nunca se utiliza para apoyar más de una página del Internet. Muchas veces no apoyan ni utilizan [versioning] que controla los contenidos ni utilizan una base de datos que controla los contenidos.

Algunos sistemas de la gestión de contenidos de código abierto se han presentado como soporte del nivel de la empresa. Plone es un sistema que se incorpora con pila de Python con el servidor de aplicación de Zope. Plone tiene muchas características contribuidas por su comunidad, pero carece muchas de las categorías del ECM tales como gerencia de documentos, gerencia de imágenes y colaboración y funciona con una base de datos de objetos propietaria en vez de una base de datos relacional estándar. Bricolage es un [WCM] basado en Perl y Postgres, pero no hace mucho más que [WCM]. Jackrabbit es la implementación de referencia de JSR-170 y también es un proyecto de Apache para apoyar y desarrollar esta implementación. Jackrabbit no fue diseñado para, ni sugieren necesariamente, usando una base de datos para almacenar metadatos. Hay algunas contribuciones para utilizar un almacén de la base de datos, pero éstas no son parte del proyecto principal.

Ningunos de los proyectos del código abierto actuales realmente comparan con las amplias capacidades o escalabilidad de los vendedores del ECM. Los proyectos de código abierto generalmente apoyan solamente la gerencia de contenidos del Web y no la conformidad, gestión de documentos o de aplicaciones de la colaboración que las empresas están buscando. No pueden tratar estas aplicaciones sin los servicios que se encuentran en un sistema comercial del ECM tal como seguridad robusta, el flujo del trabajo, la colaboración, servicios de la distribución y de la administración. Asimismo, habiendo sido incorporado de fundaciones como PHP, Perl, bases de datos propietarias o ningunas bases de datos, no pueden alcanzar la escala, anchura y distribución de los sistemas comerciales del ECM. El código abierto necesita una base nueva tecnología competir con el ECM para complementar su ventaja económica.

## **La Manera en que el Código Abierto se Junta con la Gestión de Contenidos de Empresa**

Mientras comenzamos nuestra nueva compañía, Alfresco, teníamos la ventaja de una pizarra limpia y 15 años de retrospección en el ECM. Alfresco es un sistema de la gestión de contenidos de la empresa de código abierto diseñado para tratar la oportunidad de proporcionar la gestión de contenidos de la empresa barato, conveniente, y confiable. El equipo que está levantando Alfresco es un grupo de los vendedores de la gestión de contenidos que han incorporado estos tipos de sistemas en el pasado. Elegimos una plataforma de Java para sus características de la escalabilidad y sus características par la empresa. Incorporamos muchas de las más recientes cajas de herramientas del código abierto de Java para proveernos con las capacidades requeridas para manejar servicios informativos complejos y para posibilidad de crecimiento o ampliación en ambientes de clústeres grandes. Empleamos las técnicas más últimas del diseño del interfaz de usuarios y técnicas de la implementación para simplificar el uso del sistema. También escogimos la porción del mercado donde se está gastando la mayoría del dinero, la porción del mercado de la gestión de documentos y de portales de conocimiento.

Tratando las necesidades de los usuarios que han evitado a gestión de contenidos y usado unidades de archivos compartidos, hemos rellenado una necesidad que no se ha llenado por el código abierto o de los vendedores comerciales. Para hacer esto, necesitamos tratar el paradigma de la unida de archivos de los usuarios finales completo con el acceso fuera de línea a los contenidos. Hicimos esto por agregar una implementación basad en Java del protocolo [CIFS] de Microsoft, la cual es el interfaz que usa Windows para comunicarse con discos compartidos y sincronizados de red, así dramáticamente simplificando el acceso y el uso para los usuarios finales. Proporcionamos implementaciones de portales que se podían usar inmediatamente usando el estándar nuevo de Java Server Faces para proporcionar acceso simple como Google gracias al buscador de código abierto de Lucene.

Al soportar estos interfaces, pudimos crear el depósito para proporcionar los tipos de servicios que se encuentran en un sistema comercial del ECM. Elegimos utilizar un enfoque de [Aspect Oriented Programming] para que podamos agregar funcionalidad incremental en forma de aspectos, o enchufes, que se pueden agregar independientemente de nuestra base. El marco de Spring nos proveyó nuestros interfaces del aspecto e introdujo muchos otros aspectos que nos ayudaron proveer servicios del ECM tales como control de la transacción, control de acceso basado en papeles, el hacer clústeres, interfaces de [caching and remoting]. Mirar a otros vendedores del ECM, pudimos definir servicios del Web que son potencialmente compatibles con otros sistemas también. Hemos implementado el interfaz estándar de JSR-170 para facilitar el desarrollo de nuevos aspectos en Java. Nuestro interfaz de los servicios del Web nos permite integrar con otros idiomas y ambientes tales como PHP, Perl y .NET. Otros aspectos también nos han hecho más fácil internacionalizar el sistema.

El sistema Alfresco entró la producción en octubre de 2005. Había mas de 100.000 descargas del sistema en 2005 y muchas empresas han desplegado Alfresco en el uso de la producción. Los usuarios finales han apreciado la emulación de la unida compartida de Microsoft y el uso más eficiente del portal. Miramos adelante a la comunidad de código abierto continuar descargando el sistema y proporcionar nuevos servicios del ECM en el futuro.

El sistema Alfresco se puede descargar gratuitamente en [www.alfresco.org](http://www.alfresco.org)