

Gestión de Contenido Organizacional (ECM)

Alejandra López, Julio Moreyra

Facultad de Ingeniería – Sede Trelew, Universidad Nacional de la Patagonia San Juan Bosco
Belgrano y Roca – Trelew – Chubut – Argentina
Malopez77@gmail.com, Julio.moreyra@gmail.com

Abstract. Enterprise Content Management makes reference to the handling of organizational content through printed or electronic means. It comprises the strategies, methods, and tools used to capture, handle, protect, preserve, and submit content and documents related to organizational processes.

The ECM is employed by several organizations nowadays since it provides help in the organizational process and thanks to this, it increases the company's profitability by means of improvement of the aspects related with the organization and management of documentation.

We studied the general aspects of ECM and three different products are analyzed: Alfresco, Nuxeo, and SharePoint as a ground for the comprehension of this methodology, its characteristics, scopes, benefits, and suitability.

A comparison is drawn between them based on the different perspectives: cost, community support, technology, options of execution environment, etc.

Keywords: Enterprise Content Management. Information lifecycle. Workflows.

1 Introducción

La gestión de contenido organizacional (ECM), está constituida por estrategias, métodos y herramientas usadas para capturar, gestionar, almacenar, preservar y distribuir contenido y documentos relacionados con los procesos organizacionales. Las herramientas ECM y sus estrategias permiten la gestión de la información no estructurada de la organización, donde quiera que la información se encuentre almacenada. Además, esta información debe poder ser utilizado de forma tal de alcanzar los objetivos de la organización. Son centrales a esta estrategia las herramientas y tecnologías ECM, que manejan el ciclo de vida completo de la información.

Hay cuatro áreas primarias, en las que ECM y los contenidos son fundamentales para el éxito de la organización: Cumplimiento de normas, Colaboración, Continuidad y Costos.

Cumplimiento de normas: Existen en EE.UU. regulaciones como Sarbanes-Oxley¹, HIPAA², DoD 5015.2³ entre otros, que establecen formatos y requerimientos para la gestión de registros electrónicos de documentación. En nuestro país no existen este marco regulatorio.

Debido a normativas cada vez más estrictas y con el objetivo de mostrar niveles de transparencia más altos, las organizaciones de todo el mundo recurren a tecnologías y software de gestión de los archivos electrónicos. Los costos asociados para cumplimentar una regulación, como Sarbanes-Oxley o HIPAA, son mayores e imponen nuevas exigencias en la gestión de registros. Para ayudar a limitar los riesgos y costos deben desarrollarse estrategias ECM proactivas dentro de áreas claves, tales como gestión de registros y procesos de negocios, asegurando que las prácticas apropiadas del negocio sean seguidas y el contenido sea convenientemente capturado, almacenado, gestionado y descartado en el momento oportuno y legal dentro de su ciclo de vida. Las herramientas ECM, cuando se utilizan correctamente, pueden ayudar a reducir los costos generales en el cumplimiento de normas de la organización.

Colaboración: esta permite que las personas con áreas de conocimiento complementarias o solapadas puedan crear mejores resultados de forma más expeditiva que en el pasado. Con las herramientas colaborativas actuales, las unidades de negocios y equipos pueden trabajar juntos en cualquier momento, ya sea desde oficinas contiguas o en cualquier parte del mundo. Estas tecnologías hacen posible que se puedan contemplar objetivos operacionales como la disminución de tiempos, la racionalización o coordinación de procesos, el recorte de costos y mejoras en el tiempo de comercialización. No obstante, en el uso de herramientas colaborativas, se debe tener presente la gestión de registros, captura de conocimiento y requisitos de

¹http://en.wikipedia.org/wiki/Sarbanes%E2%80%93Oxley_Act

²http://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act

³http://en.wikipedia.org/wiki/Design_Criteria_Standard_for_Electronic_Records_Management_Software_Applications

cumplimiento de normas (por ejemplo, para una industria todas las comunicaciones con el cliente deben ser guardadas).

Costos: aunque es difícil identificar el retorno de inversión directamente, no es imposible ver el impacto de un proceso mejorado de negocios. Las herramientas ECM pueden hacer que las organizaciones sean más eficientes y reduzcan sus costos, al organizar la información para su posterior recuperación, uso y por último descarte.

Continuidad: a menudo mencionado como recuperación de desastres (*disasterrecovery*), el planeamiento de continuidad de negocios es una estrategia para asegurar que las operaciones prosigan luego de una interrupción natural o provocada por el hombre. Teniendo en cuenta que los documentos electrónicos son el alma de la mayoría de los negocios, ECM juega un rol clave en la continuidad. Las tecnologías ECM permiten la creación de repositorios centralizados donde reside toda la información vital para la organización.

Cuando se utilizan metodologías tradicionales, en las que los sistemas son desarrollados ad-hoc para resolver cuestiones puntuales y/o brindar servicios que aporten valor agregado sobre información existente, se observa que cada desarrollo es llevado a cabo de manera independiente y luego resulta difícil, costoso o casi imposible integrar los distintos sistemas desarrollados. Así mismo, es complicado y laborioso, atacar los distintos aspectos que posee la información: publicación en la web, mantenimiento de registros históricos, auditoría, etc.

Surge la necesidad y conveniencia de utilizar otras soluciones en las que la captura de información, su clasificación, registro, flujo, y disposición final sea brindada como un servicio sobre el que se pueden construir todas las aplicaciones. [1]

En este trabajo nos propusimos estudiar el campo de los sistemas de Gestión de Contenido Organizacional y su aplicación en el desarrollo de sistemas.

Los beneficios que brindan estos sistemas y la experiencia de haber utilizado metodologías tradicionales, en la que los sistemas son desarrollados ad-hoc para resolver cuestiones puntuales y/o brindar servicios de valor agregado sobre información existente, y en la que cada desarrollo es llevado de manera independiente resultando su integración muy difícil y costosa, han sido el principal factor de motivación.

2 Necesidad y Oportunidad

Hay numerosos factores que llevan a las organizaciones a adoptar las soluciones ECM, por ejemplo el aumento de la eficiencia, la mejora en el control de la información y la reducción de los costos generales en su gestión. Las aplicaciones ECM facilitan el acceso a los registros a través de palabras claves y búsquedas de texto completo, brindando así a los usuarios la información necesaria directamente desde sus puestos de trabajo en segundos en lugar de acceder a múltiples aplicaciones o búsqueda manual por medio de registros en papel.

Los sistemas ECM pueden reducir las necesidades de almacenamiento, papel y envío de correspondencia, haciendo más eficientes el trabajo de sus usuarios y resultando en mejores decisiones para la organización, lo que deriva en una reducción de costos del manejo de la información.

Al comienzo hablamos de las estrategias, métodos y herramientas para ECM. Particularmente, nos concentraremos en un software gestor de contenido organizacional.

Un gestor de contenidos organizacional es una herramienta que permite la gestión de grandes cantidades de información almacenadas en forma de documentos.

La combinación de este tipo de bibliotecas de documentos con índices almacenados en una base de datos permite el acceso rápido a la información mediante diversos métodos. Dicha información está contenida en los documentos que generalmente se encuentran comprimidos y que, además de texto, pueden contener cualquier otro tipo de documentos multimedia (imágenes, videos, etc.).

Un gestor de contenidos organizacional posibilita compartir la información contenida en los documentos que son creados, editados y borrados por sus usuarios. Por consiguiente debe proveer de mecanismos que posibiliten esta colaboración y todo lo que ello conlleva: organización del repositorio de documentos, gestión de los usuarios y sus permisos para el acceso a los documentos y para la modificación de estos, control de versiones de documentos, búsquedas, etc. [2]

El gestor de contenidos organizacional, además de estas características básicas puede ofrecer otras como: notificaciones a los usuarios, reglas de publicación de documentos, mecanismos avanzados de creación de documentos a partir de plantillas, etc.

Actualmente las tecnologías que componen un ECM son descendientes de los sistemas EDMS ⁴(*Electronic Document Management Systems*) de fines de los 80 y principios de los 90. Estos eran productos autónomos (*standalone*) que proveían funcionalidad en una de cuatro áreas: captura de imágenes (*imaging*), flujos de trabajo (*workflow*), gestión documental (*documentmanagement*) y registros en discos ópticos (COLD/ERM).

La primera implementación típica incluía un sistema de captura de imágenes y flujos de trabajo a pequeña escala, posiblemente para un único departamento, con el objetivo de mejorar un proceso que requería mucho papeleo (*paper-intensive*) y el traslado hacia la mítica “oficina sin papel”.

Estos primeros sistemas autónomos fueron diseñados para ahorrar tiempo y/o mejorar el acceso a la información reduciendo las necesidades de manejo y almacenamiento de papel, y por consiguiente también se vio reducida la pérdida de documentos y se facilitó el acceso online a la información, que previamente solo estaba disponible en papel, microfilm o microfichas. Las trazas de auditoría de estos sistemas mejoraron la seguridad de los documentos y proveyeron métricas que ayudaron a medir la productividad e identificar la eficiencia.

A fines de los 90, estas tecnologías captaron la atención a las organizaciones que necesitaban soluciones enfocadas y tácticas para tratar problemas claramente definidos.

Con el transcurrir del tiempo, las organizaciones obtuvieron áreas de productividad con estas tecnologías, y quedó claro que las distintas categorías de productos EDMS eran complementarias. Debido a ello, se consideró necesaria y conveniente una amplia adopción de estos productos en toda la organización. La industria ofrecía múltiples soluciones autónomas EDMS, con poca o ninguna integración. Pero a

⁴<https://en.wikipedia.org/wiki/Edms>

finales de los 90, la integración se incrementó. A partir del año 2001, la industria comenzó a utilizar el término *Enterprise Content Management* para referirse a aquellas soluciones integradas.

En el año 2006, Microsoft con su producto Sharepoint y Oracle Corporation con Oracle Content Management se unieron a líderes establecidos de la industria tales como EMC Documentum, Laserfiche y Open Text, para ofrecer productos en el segmento *entry-level* de ECM.

Existen además productos open source, tales como Alfresco, LogicalDoc, Sense/net, eZpublish, KnowledgeTree, Jumper 2.0, Nuxeo y Plone.

En la actualidad, las organizaciones pueden desplegar un solo sistema ECM que flexiblemente gestiona la información de todos sus departamentos. Combinando ECM con soluciones de BI⁵, generan sistemas EIM (enterpriseinformationmanagement), para la gestión de la información empresarial, tanto estructurada como no estructurada.[3]

3 Análisis y Comparación

3.1 Alfresco

Es un sistema de gestión de contenido organizacional gratuito, para sistemas operativos MS Windows y Unix-like. Se ofrece en tres versiones:

CommunityEdition: como software gratuito con licencia LGPL, de código abierto y que cumple con estándares abiertos.

EnterpriseEdition: como software propietario y comercialmente licenciado de código abierto y que cumple con estándares abiertos.

In the Cloud: es la versión SaaS⁶ de Alfresco.

Incluye un repositorio de contenido, un framework de portal web para gestionar y utilizar contenido estándar de portal, una interfaz CIFS⁷ que provee compatibilidad de sistema de archivos con sistemas operativos Windows y Unix-like, un sistema de gestión de contenido web capaz de virtualizar aplicaciones web y contenido estático vía Apache Tomcat, indexación de texto completo de contenidos vía Lucene, y flujos de trabajo con el motor Activiti.

En el núcleo de un sistema Alfresco hay un repositorio que tiene el soporte de un servidor que almacena contenido, metadatos, asociaciones e índices de texto completo. Las interfaces de programación soportan múltiples lenguajes y protocolos con los que los desarrolladores pueden crear aplicaciones personalizadas y soluciones. Las aplicaciones *out-of-the-box* proveen soluciones de gestión documental, colaboración, gestión de registros y archivo, y gestión de contenido web.

Alfresco es en la actualidad la compañía más grande de gestión de contenido open source del mundo, con más de 7 millones de usuarios que utilizan versiones Enterprise, Cloud, Mobile y Community para gestionar 4.000 millones de documentos, y se utiliza en 182 países. [4][5]

⁵Business Intelligence

⁶ Software as a Service

⁷ Common Internet File System

3.2 Nuxeo

NuxeoEnterprisePlataform (Nuxeo EP) es una plataforma de sistemas de gestión de contenidos de código abierto usados por arquitectos y desarrolladores para construir, desplegar y ejecutar aplicaciones de negocios centradas en contenido. Puede ejecutarse en ambientes Windows y Unix-like.

Provee una plataforma de código abierto basada en java que es modular y extensible para el desarrollo de software ECM y que incluye un conjunto de módulos empaquetados para gestión documental, colaboración, gestión de activos digitales y gestión de casos. Entre otros, provee estos beneficios:

- Extensibilidad: arquitectura altamente flexible basada en sistemas de componentes y orientadas a servicios
- Escalabilidad: tiene buen escalamiento adaptándose a las necesidades del proyecto.
- Adaptabilidad: con Nuxeo Studio, un ambiente de personalización y configuración
- Desarrollo rápido de aplicaciones (*RAD*): manejado por una plataforma modular con componentes reutilizables.
- Código abierto, estándares abiertos: significa un compromiso a la interoperabilidad y la innovación manejado por la comunidad. [6]

3.3 SharePoint

Microsoft SharePoint es una plataforma de aplicaciones Web desarrollada por Microsoft. Lanzado por primera vez en 2001, SharePoint se ha asociado históricamente con la gestión de contenidos de la Intranet y gestión de documentos, pero con el transcurso del tiempo hoy ofrece capacidades mucho más amplias.

SharePoint está construido sobre Microsoft .Net, ASP.Net, Internet Information Server y SQL Server. Todas las plataformas deben correr en un servidor con Windows 2008 o superior de 64 bits.

Existen tres versiones:

- SharePoint Foundation
- SharePoint Standard
- SharePoint Enterprise [7]

4 Comparación de Características

Luego de analizadas en forma no exhaustiva las características de los software ECM, es posible realizar una comparativa desde el punto de vista de su forma de uso y para qué fueron concebidos.

Las alternativas open source a priori parecen no tener costos asociados a su despliegue. Sin embargo, requieren personal de IT altamente formado o bien la contratación de servicios de consultoría y soporte para lograr una implementación exitosa.

SharePoint: es un producto completo que dependiendo del tipo de licencia puede ser gratuito (Foundation) o pago, para llegar a las características empresariales (Server Enterprise) con incorporación de herramientas de BI y un ambiente totalmente integrado en el universo Microsoft. Como punto a favor: si la organización ya ha

invertido en tecnologías Microsoft, incorporar SharePoint en cualquiera de sus versiones es una transición de bajo impacto en el ambiente TIC y en los usuarios. Usa otros productos Microsoft como infraestructura (Windows Server, SQL Server, IIS) por lo que todo el conocimiento de estas plataformas es aplicable para el despliegue de SharePoint. Se ofrece un completo ambiente de desarrollo (VS 2010) en el que las aplicaciones SharePoint son reconocidas y desplegadas fácilmente. Como aspectos negativos podemos citar:

- Si bien existe una versión gratuita su funcionalidad es limitada y características como por ejemplo el modelado de contenido, vistas previas de resultado de búsquedas y tableros de control del usuario solo están disponibles en la versión Enterprise.[8]
- Su principal fortaleza es también una debilidad ya que si no se ha invertido en software de back-end Microsoft el despliegue de cualquiera de las versiones de SharePoint requerirá una inversión en los productos de infraestructura y también una atadura a un proveedor de software en particular (vendorlock in).
- Cada cliente que accede a un servidor SharePoint debe tener una licencia de acceso (CAL⁸) por lo que hay que sumarlo al costo total (TCO⁹).

Nuxeo: en pocas palabras, es una plataforma de desarrollo más que un producto terminado. Su disponibilidad de software open source con todas las características hace viable una evaluación completa. Permite independizarse del ambiente de ejecución y no tiene ataduras con ningún tipo de software de infraestructura. Su modelo de despliegue es adecuado cuando desea desarrollarse aplicaciones de ECM verticales. En el modelo de suscripción pago no se pagan licencias de acceso de clientes, y se brinda soporte técnico, mantenimiento de software y herramientas de personalización. Como aspectos negativos podemos citar:

- Su modelo de licenciamiento obliga a pagar por las herramientas de desarrollo o de otra forma hay que recurrir a la edición de archivos XML para configurar distintos aspectos de la aplicación: branding, workflows, interfaces de usuario, etc.
- La comunidad de usuarios y desarrolladores no es tan extendida como en el caso de Alfresco.
- No tiene herramientas de BI out-of-the-box.

Alfresco: al igual que Nuxeo, al ser software open source permite una evaluación completa de sus características. Posee una interfaz web de usuario rica y simple, con mucha potencia y capacidades de extensión a través de *gadgets* que se incorporan en los tableros de control del usuario. También permite independizarse del ambiente de ejecución. La comunidad Alfresco es muy amplia y activa brindando un gran número de *addons* gratuitos o pagos. Es una aplicación madura con una gran base de usuarios, y es el software ECM open source más utilizado en el mundo. Su modo de licenciamiento *on-premise* no requiere de licencias de acceso al cliente y ofrece soporte técnico y mantenimiento. Como aspectos negativos podemos citar:

- No tiene herramientas de BI out-of-the-box.
- No posee herramientas de desarrollo integradas gratuitas. Existen addons con funcionalidad limitada y hay que recurrir a la edición de archivos XML para otras configuraciones.

⁸ Client Access License

⁹ Total Cost of Ownership

En la tabla 1 se resumen las características que a nuestro criterio pueden resultar de interés a la hora de la toma de decisiones sobre la elección de una herramienta sobre otra para un proyecto en particular.

Tabla 1. Comparación de características

Característica	Alfresco	Nuxeo	SharePoint
Open source	Si	Si	No
Versión gratuita	Si	Si	Si ¹⁰
Vendorlockin	No	No	Si
Soporte de BI integrado	No	No	Si
Licenciamiento por cantidad de clientes	No	No	Si
Gestión del ciclo de vida de la información	Si	Si	Si
Cumplimiento de regulaciones de gestión de registros (HIPAA, Sarbanes-Oxley, e-discovery)	Si	No	Si
Versión <i>in thecloud</i>	Si	Si	Si
Herramientas de desarrollo gratuitas	Si ¹¹	No	Si
Herramientas de desarrollo provistas por el fabricante	No	Si	Si
Herramientas de administración completas e integradas (backup/restore, gestión de componentes, aplicaciones, etc)	No	No	Si
Trabajar desconectado (cliente de sincronización offline)	Si	Si	Si

5 Desarrollo sobre Alfresco

Procedimos a desarrollar una aplicación pequeña sobre Alfresco, con el objetivo de evaluar el nivel de dificultad que esto conlleva. Elegimos un organismo no informatizado dentro de nuestro ambiente laboral del Poder Judicial como escenario para la implementación de un sistema informático. El organismo elegido es la Dirección de Mediación del Poder Judicial de la Provincia del Chubut, que es la encargada de organizar el Registro Provincial de Mediadores y el Servicio Público de Mediación. Por razones de espacio, en este trabajo no se describe la problemática con mayor detalle, sólo se quiere mencionar el resultado de la experiencia.

Habiendo definido el modelo de datos de acuerdo al dominio de aplicación, observamos que rápidamente podemos dar soporte a múltiples requerimientos con las capacidades out-of-the box: subir contenido al repositorio, realizar búsquedas personalizadas, streaming de contenido, gestionar usuarios, permisos de acceso, diseñar micrositios con paneles personalizados, acceder y actualizar contenido mediante protocolo CIFS, asignar y modificar metadatos al contenido, etc.

¹⁰ Con funcionalidad limitada y requiere una licencia de Windows Server.

¹¹ Addons de terceras partes, con capacidades limitadas

Para la implementación de la aplicación mencionada anteriormente elegimos herramientas gratuitas y open source. La dificultad que esto conlleva es la obtener un ambiente productivo con rapidez, teniendo que ensamblarse parte por parte. También es dificultosa la obtención de documentación de algunas herramientas, y sólo se dispone de los foros de usuarios. Sin embargo, la experiencia es altamente favorable y nos alienta a utilizar el modelo de gestión de contenido para el desarrollo para futuras aplicaciones.

6 Conclusiones

Uno de los problemas que todas las organizaciones enfrentan en la incorporación de TIC es la dispersión de los contenidos digitales y la gestión del ciclo de vida desde su creación o captura hasta la disposición final. También deben tenerse en cuenta cuestiones como la correcta catalogación, almacenamiento, preservación, distribución, su búsqueda y el seguimiento de políticas de acceso a la información. Esto representa un desafío en cada nuevo sistema a desarrollar. Así mismo, estos sistemas deben integrarse y colaborar con otros existentes.

Es por ello y tal como indicamos al inicio del presente trabajo que la gestión del contenido organizacional es una necesidad, constituida no sólo por herramientas de gestión de contenido, sino también por estrategias y métodos. Así podremos hacer frente a la gestión de la información no estructurada, donde quiera que la información se encuentre.

Habiendo analizado tres productos testigo de gestión ECM tanto en el campo comercial como open source hemos podido apreciar sus potencialidades para ser utilizados como infraestructura de base en la creación de nuevos sistemas.

Las herramientas analizadas cumplen con una gran gama de requerimientos funcionales que se presentan en la gestión de contenido organizacional. Out-of-the box proveen las siguientes características, entre otras:

- Gestión del ciclo de vida de la información
- Cumplimiento de normas de gestión de registros
- Acceso desde dispositivos móviles
- Sincronización con clientes desconectados
- Selección de implementación on-premise/in thecloud
- Cumplimiento de políticas organizacionales de acceso a la información

La elección del producto dependerá de:

1. la infraestructura de software back-end existente
2. el recurso humano disponible de IT, su formación y experiencia
3. el recurso financiero asignado a la implementación del proyecto

Para el caso mostrado, la elección recayó en Alfresco basados en un conocimiento previo del software, el deseo de utilizar herramientas open source y la idea general que Alfresco es un producto más maduro y completo que la otra alternativa open source (Nuxeo). Sin embargo, cualquiera de ellos se adecuaba al propósito planteado.

Con la experiencia de haber utilizado con anterioridad metodologías tradicionales en las que los sistemas son desarrollados ad-hoc, reconocemos la importancia de este tipo de herramientas que brindan soluciones robustas para la gestión de información, su clasificación, registro, flujo, y disposición final. Que esto sea brindado

como un servicio sobre el que se pueden construir todas las aplicaciones resulta ser una solución muy superadora.

7 Referencias

- [1] What is ECM? What is Enterprise Content Management?, <http://www.aiim.org/What-is-ECM-Enterprise-Content-Management.aspx>
- [2] Manual de Usuario de Alfresco,
https://documenta.ugr.es:8443/alfresco/d/d/workspace/SpacesStore/b155c802-f825-4e49-9ab6-256625c7d2be/ManualAlfresco_CSIRC_v1.0.pdf
- [3] Enterprise Content Management,
http://en.wikipedia.org/wiki/Enterprise_Content_Management
- [4] Alfresco (software), [http://en.wikipedia.org/wiki/Alfresco_\(software\)](http://en.wikipedia.org/wiki/Alfresco_(software))
- [5] Alfresco Accelerates Growth and Adds Over 500 Enterprise Customers,
<http://www.alfresco.com/news/press-releases/alfresco-accelerates-growth-and-adds-over-500-enterprise-customers>
- [6] Enterprise Content Management Platform for Business Apps,
<http://www.nuxeo.com/en/products/content-management-platform>
- [7] Microsoft SharePoint, http://en.wikipedia.org/wiki/Microsoft_SharePoint
- [8] Khamis, N. SharePoint 2010 Standard vs Enterprise vs Foundation edition features .
<http://www.khamis.net/blog/Lists/Posts/Post.aspx?ID=61>