

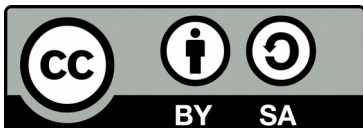
**Software libre para la
gestión documental**

Módulo 4

Enterprise Content Management (ECM) libres

Guillermo Castellano





Enterprise Content Management (ECM) libres

Guillermo Castellano Casas

Septiembre 2017

Maquetado por Guillermo Castellano Casas con LibreOffice 5.0 en Linux Mint 17.3.

Fuente Alegreya Sans de Juan Pablo del Peral.

Imagen de portada de [Pexels](https://www.pexels.com/).

Infografías diseñadas por Ana Amelia Patiño Esteo con Inkscape en Ubuntu 16.04.

Puedes compartir, es decir, copiar y distribuir el material en cualquier medio o formato y adaptar, es decir, mezclar, transformar y crear a partir del material. Las condiciones para esto son reconocer adecuadamente la autoría, proporcionando un enlace a la licencia e indicando si se han realizado cambios (haciéndolo de cualquier forma razonable, sin sugerir que se tiene el apoyo del licenciador o lo recibe por el uso que hace), y compartiendo con la misma licencia si se mezcla, transforma o crea a partir del material.

Enterprise Content Management (ECM) libres

1. Qué es un Enterprise Content Management

Antes de analizar algunos de los ECM (siglas en inglés de Enterprise Content Management) libres y de código abierto de los que disponemos en el mercado, es importante aclarar a qué nos referimos exactamente por ECM y si es esta denominación significa lo mismo que software de gestión documental.

En el mercado del software de gestión de contenidos nos encontramos con denominaciones muy diversas:

1. Document Management System (DMS).
2. Electronic Document Management System (EDMS).
3. Records Management System (RMS).
4. Electronic Records and Document Management System (ERDMS).
5. Digital Asset Management (DAM).
6. Enterprise Reports Management (ERM).
7. Content Management System (CMS).
8. Enterprise Content Management (ECM).
9. Web Content Management (WCM).
10. Document Control System (DCS).

La existencia de múltiples siglas para referirse a productos con funcionalidades muy similares o, en muchos casos, idénticas, solamente contribuye a crear confusión. Una búsqueda rápida en internet sobre algunas de las principales soluciones de este mercado verifica que no existe un consenso a la hora de cómo denominarlas, por mucho que en un plano estrictamente teórico existan matices que justifiquen diferentes denominaciones.

A grandes rasgos, las herramientas informáticas de gestión de contenidos se dividen en tres categorías, según la naturaleza del contenido que procesan:

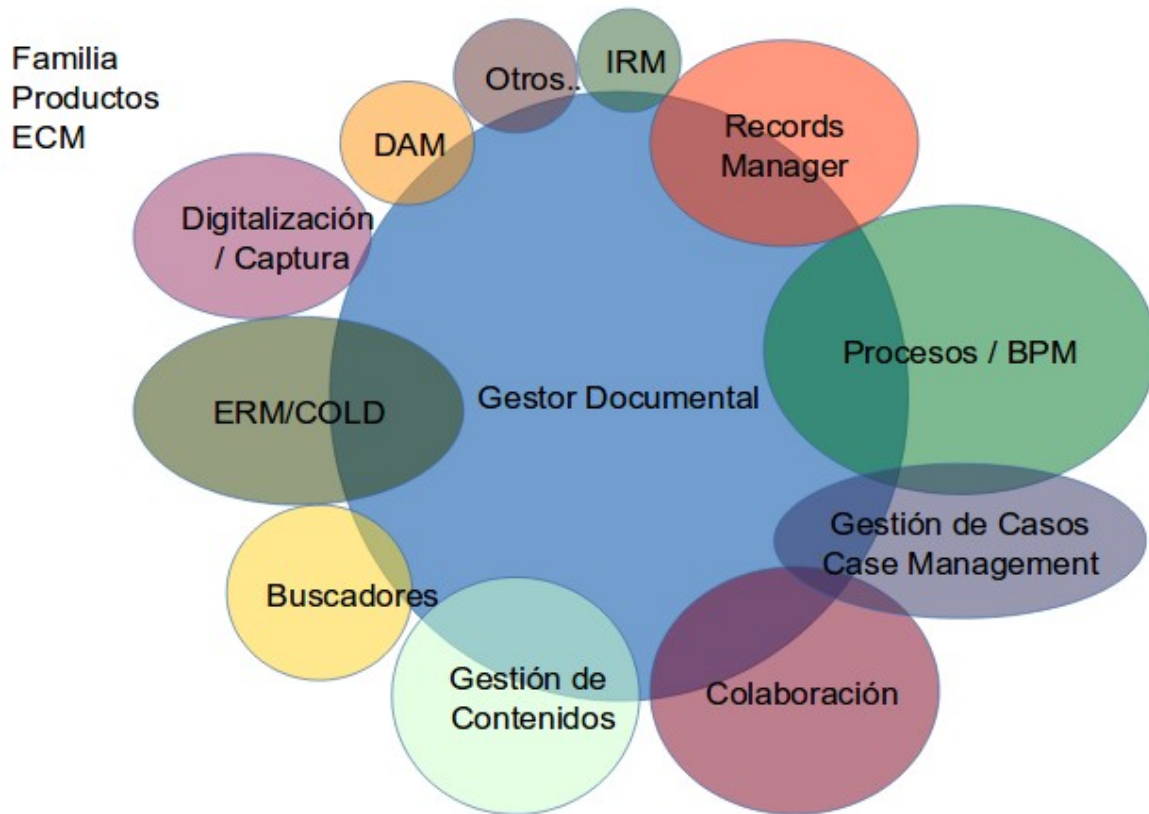
1. Software de gestión de documentos y registros.
2. Software de gestión de contenidos web.
3. Software de gestión de contenidos empresariales.

Los ECM entran dentro de la tercera categoría, ya que ambicionan gestionar todo el contenido incluido por la empresa, incluyendo tanto los documentos electrónicos y registros como los contenidos web. ¿Es correcto ECM sinónimo de software de gestión de documentos electrónicos? ¿No sería más correcto referirnos al software de gestión de documentos electrónicos como ERDMS? La respuesta depende de lo que consideremos documento.

Para la archivística, los documentos son la evidencia de la actuación, única o secuencial, del sujeto productor, que puede ser un organismo público o privado o una persona. Tradicionalmente, esta definición se aplicaba únicamente al documento en papel, pero la extensión de las tecnologías de la información a los procesos administrativos ha obligado a ampliar lo que entendemos por documento a otros soportes de información. Las normas internacionales sobre gestión documental reflejan esta nueva realidad. Así, la familia UNE-ISO 30300:2011 define el documento como “información u objeto registrado que puede ser tratado como una unidad”, mientras que la nue-

Módulo 4. Enterprise Content Management (ECM) libres

va UNE-ISO 15489:2016 lo define como una “evidencia de actividad funcional”. En otras palabras, un correo electrónico, un registro informático o un contenido web también son documentos, en la medida en que se trata de evidencias producidas por las organizaciones sobre su actividad, por lo que un ECM puede ser equiparado perfectamente a un software de gestión documental¹.



Lean Proposal Canvas.

Fuente: Kleimer.

A nivel técnico, un ECM es una plataforma formada por funcionalidades de otras categorías de productos más específicas. Su núcleo comprende las funcionalidades típicas de un producto de gestión documental, a saber:

1. Clasificación de los documentos en varios niveles (fondo, serie, expediente).
2. Definición de tipos documentales.
3. Gestión de metadatos.
4. Control de versiones de los documentos.
5. Acceso controlado a la documentación (ACL).

A estas funcionalidades se suelen agregar otras, entre las que destacaría las siguientes:

1. Motor de búsqueda.
2. Digitalización de documentos en papel.

¹ Obsérvese que esta definición de documento engloba tanto el “document” como el “record” de la tradición archivística anglosajona, que diferencia entre el documento y el registro. De hecho, en castellano traducimos la expresión “records management” como “gestión documental”, no como “gestión de registros”.

Módulo 4. Enterprise Content Management (ECM) libres

3. Reconocimiento Óptico de Caracteres.
4. Creación y gestión de flujos de trabajo o workflows (BPM).
5. Colaboración.
6. Gestión de contenidos audiovisuales.
7. Gestión de contenidos web,

Como vemos, este tipo de programas reúnen las funcionalidades suficientes para asegurar que los documentos electrónicos se mantengan auténticos, íntegros, fiables y usables a lo largo de todo su ciclo vital, que es de lo que se ocupan la gestión documental y la archivística. Uno de los elementos más importantes de los ECM para lograr esto son los metadatos. Estos datos que describen otros datos nos permiten:

1. Recuperar la información: Los metadatos permiten recuperar y filtrar la información por criterios más variados que la clasificación y el título del documento.
2. Automatizar procesos: Gracias a los metadatos, podemos automatizar procesos como la descripción o la conservación.
3. Asegurar la trazabilidad: Los metadatos pueden guardar información sobre eventos como la lectura, escritura y borrado de un documento.

Existe una gran variedad de ECM de código abierto que está pisando fuerte en el mercado de este tipo de aplicaciones, entre los que destacan Alfresco, Nuxeo, OpenKM, OpenProdac y, prácticamente desconocido en el ámbito español y latinoamericano, SeedDMS (del que no hablaremos en este curso).

Una de las noticias que nos dejó 2016 es la adquisición de Documentum por OpenText. Aunque OpenText augura un futuro prometedor a la antigua plataforma de Dell EMC, continúa sin dar muchos detalles sobre el roadmap de Documentum. Esta incertidumbre no sienta nada bien a un producto que lleva acumulados 4 años seguidos de disminución de los beneficios. Algunos analistas piensan que la estrategia de OpenText consistirá en rentabilizar la base de clientes de Documentum y tratar de venderles sus propios productos y sus servicios basados en la nube. John Newton, co-fundador de Documentum, predice que OpenText no invertirá en el desarrollo a largo plazo de esta plataforma y la dejará morir poco a poco para que sus usuarios actuales se pasen a su propio software. Las palabras de Micah Byrd (2017) a mediados del pasado mes de abril sobre la integración de software de OpenText en Documentum parecen confirmar estas sospechas.

En contraste con esta inquietud sobre el futuro de uno de los principales ECM privativos, en 2016 Alfresco se clasificó por primera vez como desafiante en el Cuadrante Mágico de Gartner. Una de las fortalezas de Alfresco es su reciente alianza con Amazon Web Services para ofrecer el despliegue de su plataforma en los servidores del gigante del comercio electrónico. Otros ECM libres como Nuxeo y el español OpenKM también están disponibles en la nube, permitiendo una implementación más simple y adaptada al tamaño y las necesidades de la organización.

Con todo, la gran baza de los ECM de código abierto sigue siendo que permiten a las organizaciones con menos recursos acceder a herramientas de gestión documental profesionales. Una de las tendencias de las PYMES, las administraciones públicas y las entidades del tercer sector es prestar cada vez más atención a estas plataformas no privativas y dar la espalda a soluciones caras, complicadas de implementar y difíciles de dejar atrás una vez las hemos implementado como Documentum.

Otra de las ventajas del software libre y de código abierto —que también mencionamos en el segundo módulo de este curso— es su interoperabilidad. En este módulo tendremos ocasión de explicar la arquitectura de varios ECM libres y de código abierto y ver cómo favorece la integración con otras aplicaciones, sistemas y repositorios.

Asimismo, cuando una organización decide gestionar su documentación con un ECM libre y de código abierto se asegura la independencia con respecto al proveedor. Esto le proporciona no solamente la posibilidad de cambiar a otro proveedor si no está conforme con el servicio que recibe o el precio que paga, sino también la tranquilidad

dad de que seguirá teniendo la propiedad de todo su contenido (tanto documentos binarios como metadatos) en caso de migración a otra plataforma. Qué va a ocurrir con toda la información en caso de cambio a otro programa es un aspecto que conviene tener muy presente durante los procesos de selección de software de gestión documental para evitar un divorcio conflictivo y costoso.

En el módulo anterior hemos visto que a la hora de valorar una herramienta de este tipo hay que tener en cuenta criterios técnicos, funcionales y de servicio. Por este motivo, el análisis de que realizaremos a continuación de Alfresco, Nuxeo, OpenKM y OpenProdoc se centrará en sus funcionalidades, su arquitectura y su distribución. El objeto de este módulo no es enseñar a utilizar estos programas, sino que el alumno adquiera una visión comparada de ellos y conozca las debilidades y fortalezas de cada uno, con el fin de entender qué pueden ofrecerle los ECM libres y de código abierto.

2. Alfresco

2.1. Descripción

Comenzamos nuestra lista por el ECM de código abierto más famoso de todos: Alfresco. La compañía detrás de este software fue creada en 2005 por John Newton, cofundador de Documentum, y John Powell, antiguo directivo de Business Objects, con el objetivo de sacar al mercado una alternativa a las herramientas de gestión documental privativas, que en ese momento tenían el monopolio del mercado. Nació como un producto centrado en la gestión documental pura y dura, pero ha evolucionado hacia una plataforma más compleja que incluye también gestión de contenidos web, motor de flujos de trabajo, etc.

La reputación de Alfresco se debe en parte a su buen posicionamiento en las informes de las consultoras Gartner, que en 2016 lo clasificó como desafiante en su Cuadrante Mágico, y de Forrester, para quien este ECM rivaliza con las soluciones privativas de IBM (FileNet) y de OpenText (Documentum). Hay que tener en cuenta que estos informes utilizan criterios de inclusión no técnicos, como pueden ser la facturación y el grado de internacionalización del fabricante, que excluyen las demás soluciones de código abierto, desarrolladas normalmente por compañías más pequeñas o por la propia comunidad.

Otra de las razones por las que Alfresco destaca en el mercado es su reciente alianza con Amazon Web Services para ofrecer el despliegue de su plataforma en los servidores del gigante del comercio electrónico en unos pocos minutos. Antes de que existiera esta posibilidad, muchos proveedores de hosting ya ofrecían scripts para desplegar en sus servidores instancias con Alfresco en un solo clic, lo que facilita su adopción por parte de empresas con recursos informáticos limitados (que suelen ser las destinatarias de este tipo de servicios) y reduce el coste total de la solución.

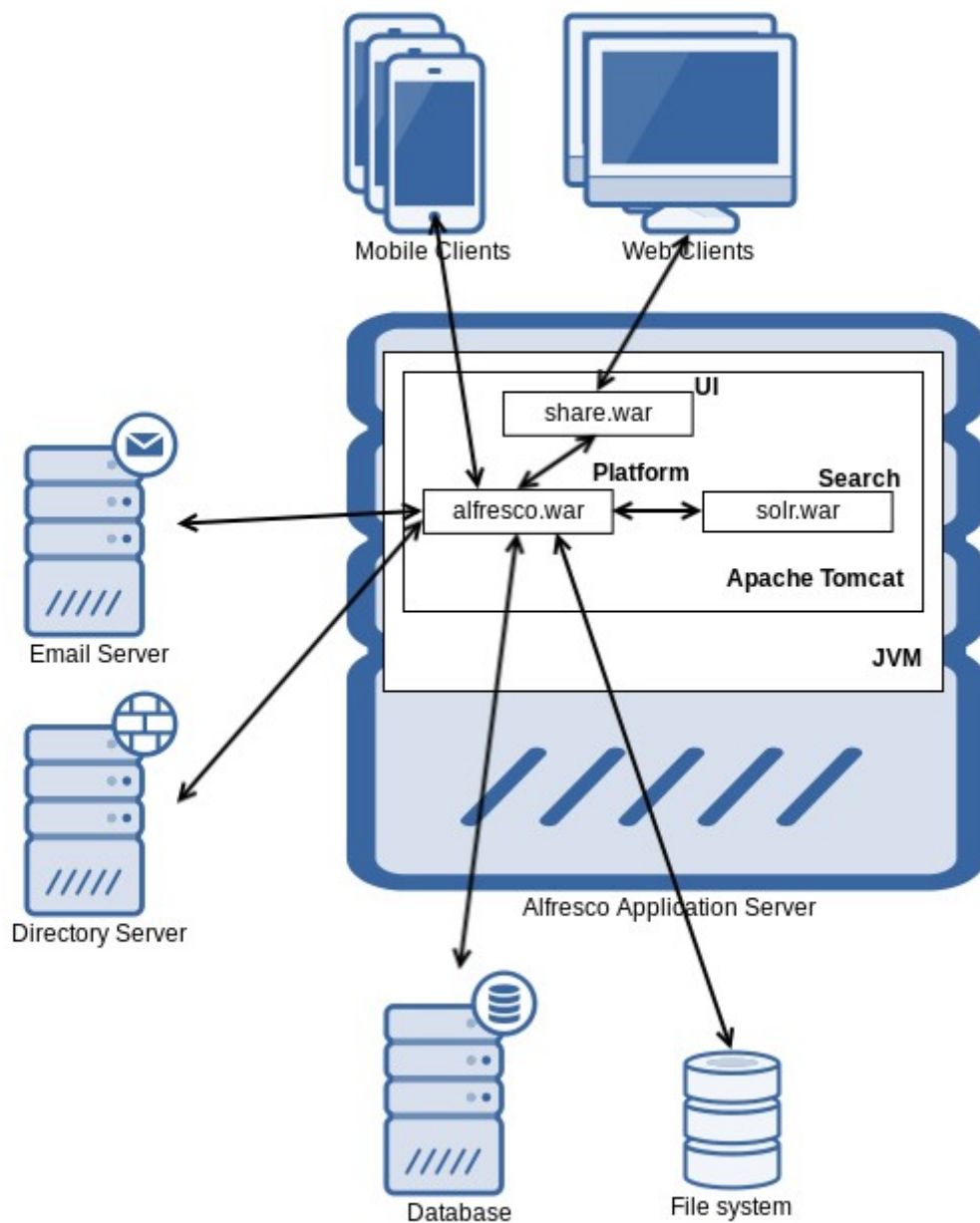
Entre sus funcionalidades, destacan las siguientes:

1. Gestión documental.
2. Integración con Activity BPM para la creación y gestión de flujos de trabajo.
3. Gestión de contenido en dispositivos móviles.
4. Creación y gestión de sitios colaborativos.
5. Gestión de contenido web.
6. Integración con escáneres y OCR para automatizar la digitalización y la captura de datos.
7. Desarrollo de aplicaciones.

2.2. Arquitectura

Alfresco se compone de tres archivos WAR o archivos de aplicación web:

1. Share, que es la interfaz de usuario.
2. Alfresco, que es la plataforma.
3. Solr, que es el buscador.



Arquitectura de Alfresco.
Fuente: *Alfresco One 5.1.3 Developer guide*

Módulo 4. Enterprise Content Management (ECM) libres

Estas aplicaciones se despliegan dentro de un servidor de aplicaciones J2EE, a las que el usuario accede mediante un navegador web. Alfresco suele funcionar con Tomcat, que es de código abierto, pero es compatible igualmente con Jboss (también libre), WebLogic y WebSphere. Este servidor va instalado dentro de una máquina virtual Java (JVM), lo que permite que el ECM se ejecute en cualquier sistema operativo que soporte una JVM: GNU/Linux, Microsoft Windows, macOS, FreeBSD, etc.

La plataforma apunta también a la base de datos (donde se almacenan los metadatos), el sistema de archivos (donde se almacenan los archivos binarios, es decir, los documentos), el servidor de correo electrónico, el directorio del servidor y todas las extensiones que queramos añadir. Dicho de otro modo, se trata del nodo donde convergen todos los elementos del sistema. Como vemos, Alfresco trabaja con una arquitectura n-niveles que permite que los diferentes elementos del ECM se distribuyan sin problema en varios servidores.

Alfresco es compatible con las principales bases de datos relacionales del mercado, tanto libres como privadas: MySQL, MariaDB, PostgreSQL, Oracle, etc. También se puede utilizar una base de datos NoSQL (por ejemplo, MongoDB) como back-end de este ECM.

Alfresco se integra con multitud de software de terceros, entre ellos programas de reconocimiento óptico de caracteres (Alfresco es compatible, por ejemplo, con el OCR de código abierto Tesseract), motores de flujos de trabajo (Alfresco es compatible con varios BPM del mercado, aunque el fabricante ha desarrollado su propio BPM, Activiti), CMS como Liferay y Magnolia, etc. También soporta CMIS.

2.3. Distribución

Alfresco se distribuye en tres versiones: una versión Community, que es totalmente libre y se distribuye bajo licencia LGPLv3, y las versiones Cloud y Content Services, que incorporan suscripción a diferentes servicios de pago, entre ellos funcionalidades recortadas a la versión Community.

Una crítica frecuente a Alfresco es que ofrece un producto de código abierto limitado como gancho para que el cliente lo pruebe y se pase a la versión de pago, que incluye todas las actualizaciones y prestaciones. La propia empresa reconoce que sacaron al mercado Alfresco como software libre para conseguir más cuota de mercado y que “el código abierto no es uno de [sus] principios rectores” (Potts, 2015). Alfresco Community está diseñado para que el futuro cliente se haga a la idea de las posibilidades de la versión profesional de Alfresco y decida pasarse a una versión de pago. El planteamiento del fabricante es que si la implementación es lo suficientemente crítica como para requerir las funcionalidades recortadas, la organización usuaria debería suscribirse a una modalidad de pago (Potts, 2012). De hecho, las aportaciones de la comunidad ya no son incorporadas al código de Alfresco Community y algunos consultores especializados en ECM de código abierto no descartan que la edición de código abierto desaparezca en el futuro (Conradt, 2017).

Éstas son algunas de las diferencias funcionales que se pueden encontrar entre las versiones de pago y la versión Community:

1. Consola de administración avanzada.
2. Configuración de clústeres mediante interfaz gráfica.
3. Capacidad de definir múltiples almacenes de contenido.
4. Cifrado de contenido.
5. Integración con software de terceros como Kofax, Amazon S3 o Outlook.
6. Módulos opcionales de Alfresco.
7. Desktop Sync (sincronización del directorio local y el repositorio de Alfresco).

En la siguiente tabla se puede ver una comparativa de los diferentes servicios de soporte incluidos en cada distribución de Alfresco.

Módulo 4. Enterprise Content Management (ECM) libres

	Alfresco Community Edition	Alfresco Cloud	Alfresco Content Services
Instalación en la nube	-	Incluido	-
Instalación en servidor de la organización	-	-	-
Escalado de bugs, parches, paquetes de servicios y procesos de prioridad	-	Incluido	Incluido
Hosting	-	Incluido	-
Soporte de la comunidad	Incluido	Incluido	Incluido
Soporte profesional	-	Incluido	Incluido
Documentación	Incluido	Incluido	Incluido

Además de estos servicios, Alfresco y su red de partners ofrecen servicios de consultoría, instalación, parametrización, formación, etc. La formación –que, como vimos en el segundo módulo, constituye uno de los aspectos más determinantes en cualquier proyecto de implementación de software de gestión documental– está dividida en módulos cuyo coste oscila entre 199 \$ y 2085 \$ cada uno. La empresa también oferta un “pasaporte” que da acceso en línea a todos los módulos durante un año por 3450 \$.

The plain truth about Alfresco's open source ethos (Potts, 2015)

In the early days, open source was attractive to the company not because it wanted help building the software, but because the license undermined the position of proprietary vendors and because they hoped to gain market share quickly by leveraging the viral nature of freely-distributable software. Being open was an attractive (and highly marketable) contrast to the extremely closed and proprietary nature of legacy ECM vendors such as EMC and Microsoft.

I think John and Paul also hoped that the open and transparent nature of open source would lend itself to developer adoption, third-party integrations and add-ons, and a partner ecosystem, which it did.

I think it is this last one—the mismatch between the original motivations to release as open source and what we as a community expect from an open source project—that causes angst. The “open source” moniker attracts people who wish the project was more like an organic open source project than it can or ever will be.

For me, personally, I accepted these as givens a long time ago—none of them bother me any more. I am taking this gift that we've been given—a highly-functional, freely-distributable ECM platform—and I'm using it to help people. I'm no longer interested in holding the company to a dogmatic standard they never intended to be held to.

2.4. Debilidades

Las principales debilidades de Alfresco son las siguientes:

1. Muchas funcionalidades centrales de este ECM no son accesibles en la edición Community y no se tiene en cuenta a la comunidad en el desarrollo de Alfresco Community.

Módulo 4. Enterprise Content Management (ECM) libres

2. La administración de Alfresco es compleja y la documentación técnica al respecto resulta poco clarificada.
3. El coste de los programas de formación ofrecidos por el fabricante es muy elevado.

2.5. Fortalezas

Las principales fortalezas de Alfresco son las siguientes:

1. Se trata de un programa modulable y que ofrece muchas posibilidades.
2. El fabricante cuenta con una red de partners oficiales muy amplia en todo el mundo.
3. Alfresco Community cuenta con una comunidad muy amplia y es fácil encontrar documentación, soporte, scripts y add-ons. Una de las organizaciones de referencia dentro de su comunidad es Order of the Bee, con decenas de miembros repartidos en todo el mundo.

3. Nuxeo

3.1. Descripción

Aunque hoy en día Alfresco es el ECM de código abierto más conocido, cuando salió al mercado en 2006 el proyecto CPS ya llevaba unos años en marcha. Nuxeo CPS fue el antecedente del Nuxeo que conocemos hoy en día y ya presentaba una de las características por las que, todavía a día de hoy, destaca esta herramienta de gestión documental: la ausencia de una versión de pago con un código diferente a la versión libre. El código del programa que utilizan los usuarios que no son de pago es exactamente el mismo que el de los usuarios que sí lo son.

Una de las limitaciones de Nuxeo CPS es que estaba escrito en Python, mientras que Alfresco estaba escrito en Java desde sus orígenes. La compañía era consciente de las limitaciones de Python con respecto al lenguaje de programación de Sun Microsystems y en 2006 anunció que iba a reescribir el código de su programa para convertirlo en un ECM de nueva generación (Humble, 2010). El resultado de este cambio de estrategia fue Nuxeo Platform, que vio la luz en 2010 y se ha convertido en uno de los principales ECM del mercado, con especial presencia en Francia y América del Norte, si bien en España se distribuyó también de la mano de Yerbabuena Software hasta que esta start-up andaluza comercializó su propio software privativo basado en Nuxeo (Athento).

Desde el punto de vista de las funcionalidades y la arquitectura, Nuxeo es similar a Alfresco, aunque presenta algunas diferencias que vamos a ver más detalladamente en el siguiente epígrafe, como el uso de Elasticsearch en vez de Solr, el funcionamiento con OSGi, etc.

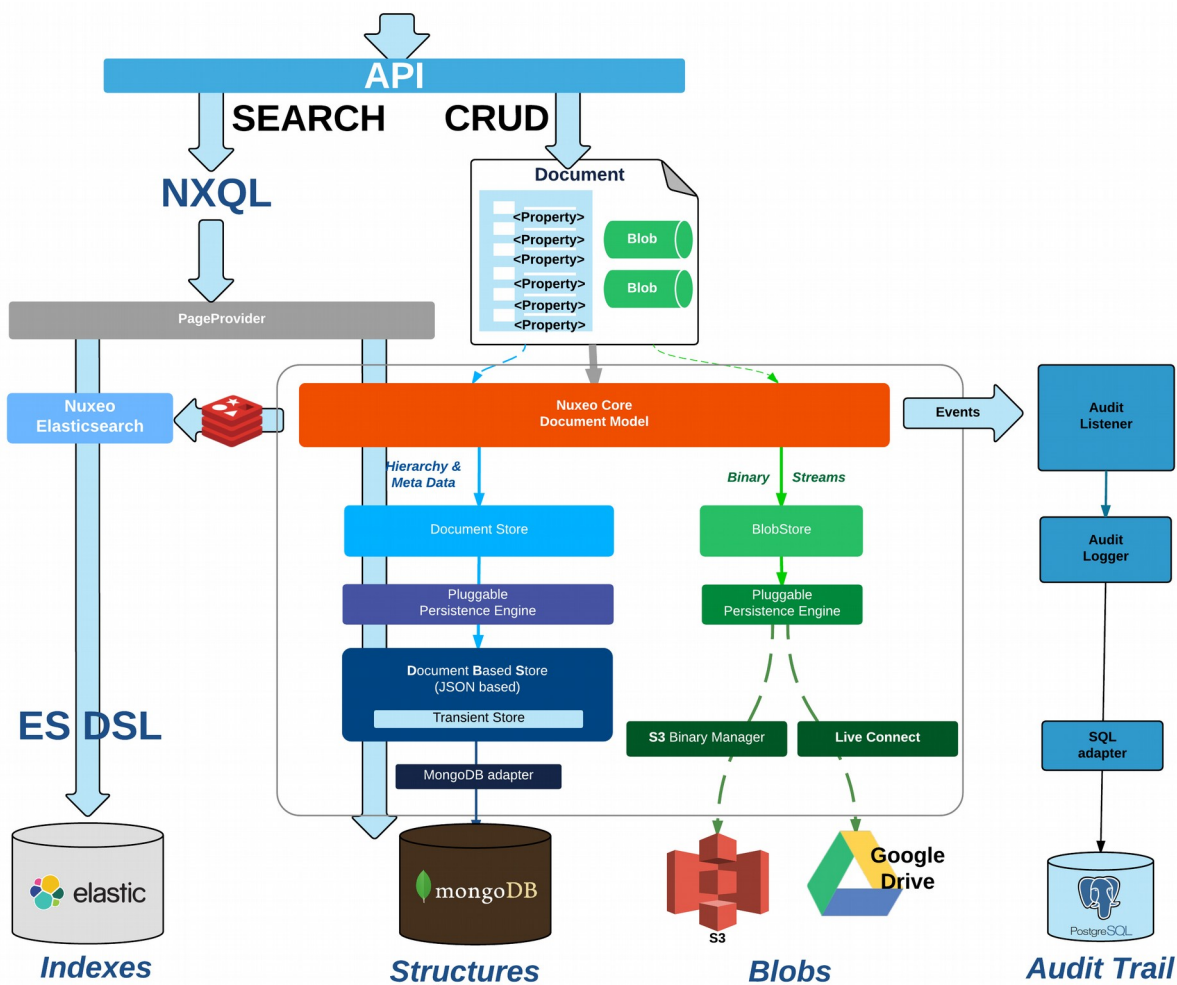
3.2. Arquitectura

Nuxeo Platform está construido en Java y está pensado para funcionar con OSGi. También es compatible con otros servidores y contenedores de aplicaciones como Tomcat, JBoss y Jetty. El cambio de Python a Java se realizó buscando una mayor compatibilidad con los estándares de código abierto, ya que, al igual que ocurre en Alfresco, el núcleo del programa está concebido solamente como un componente más de un ecosistema más amplio formado por la base de datos, el sistema de archivos, el buscador, el motor de flujos de trabajo, repositorios de terceros, etc. En concordancia con esta filosofía, Nuxeo cumple totalmente el estándar CMIS y también se puede integrar mediante API.

Nuxeo trabaja tanto con bases de datos relacionales como PostgreSQL, como con bases de datos NoSQL como MongoDB. La herramienta almacena las jerarquías y los metadatos en esta base de datos y guarda los documentos binarios en el sistema de ficheros o en repositorios de terceros como Amazon S3 o Google Drive. Para la recuperación de información, Nuxeo utiliza Elasticsearch, basado en Apache Lucene.

Esta plataforma es compatible con Nuxeo Studio, un software privativo desarrollado para facilitar la configuración y parametrización de la herramienta a los administradores con escasos conocimientos de programación.

Módulo 4. Enterprise Content Management (ECM) libres



Arquitectura de Nuxeo

Fuente: Nuxeo Documentation Center Home

3.3. Distribución

Nuxeo carece de una versión de pago diferente a la versión libre y de código abierto. La única modalidad disponible del software es Nuxeo Platform, distribuido bajo la licencia Apache v2.0 (que, recordemos, permite distribuir los productos derivados bajo licencias no libres). Incluso extensiones como Nuxeo Drive, una herramienta para sincronizar el directorio local con el repositorio de Nuxeo, son libres.

El modelo de negocio de Nuxeo se basa en vender soporte, consultoría, formación, hosting (tanto en sus propios servidores como en Amazon Web Services) y el software privativo Nuxeo Studio, que sirve para configurar y parametrizar Nuxeo Platform de manera más sencilla. La siguiente tabla resume los diferentes productos y servicios de pago ofrecidos por Nuxeo.

Módulo 4. Enterprise Content Management (ECM) libres

Nuxeo Studio	Soporte	Consultoría	Formación	Cloud hosting
Herramienta web para configurar y personalizar Nuxeo Platform a través de una interfaz gráfica	Soporte profesional por parte del equipo de Investigación y Desarrollo de Nuxeo	Puesta en marcha del proyecto	Configuración de Studio	Hosting para Nuxeo Cloud.
		Taller de arquitectura	Administración del sistema	
		Configuración	Desarrollo e integración	
		Revisión del código	REST API	
		Asesoramiento para desarrollos		
		Evaluación y mejora del rendimiento		

3.4. Debilidades

Las principales debilidades de Nuxeo son las siguientes:

1. La administración de Nuxeo Platform puede resultar compleja sin el software privativo Nuxeo Studio.
2. Está poco extendido en el ámbito español y latinoamericano en comparación a las demás soluciones de código abierto que estamos estudiando.

3.5. Fortalezas

Las principales fortalezas de Nuxeo son las siguientes:

1. No existe una versión de pago con funcionalidades recortadas de la versión gratuita.
2. Se trata de un programa modulable y que ofrece muchas posibilidades.
3. Nuxeo proporciona a los usuarios documentación técnica muy bien estructurada y la posibilidad de seguir itinerarios de aprendizaje.
4. Arquitectura basada en OSGi, más flexible que otros frameworks.

4. OpenKM

4.1. Descripción

OpenKM es un ECM desarrollado por la compañía OpenKM Knowledge Management SL, con sede en Palma de Mallorca. El proyecto comenzó en 2006 con el objetivo de hacer accesible a las pequeñas y medianas empresas la gestión de documentos electrónicos, reservada en aquella época a las grandes empresas, que eran las únicas que podían asumir los altos costes de implementación de las soluciones privativas.

Este ECM reúne todas las funcionalidades que se esperan de un programa de este tipo, incluidas la creación y gestión de flujos de trabajo (OpenKM se integra con jBPM y Bonita BPM) y la extracción de caracteres mediante OCR (OpenKM se integra de forma nativa con varios programas de reconocimiento óptico de caracteres, entre ellos Tesseract y Cuneiform, que son de código abierto, y también se puede integrar con OCR privativos como Nuance). OpenKM pone su acento en la reutilización de los documentos en tanto que activos de información y de negocio y proporciona al usuario muchas herramientas para describir y recuperar los documentos, desde las clásicas búsquedas en base al cuadro de clasificación o el nombre del documento, a búsquedas por metadatos, por texto libre, etc. El usuario también puede ver sus búsquedas frecuentes, suscribirse a carpetas, documentos y registros y enviar no-

Módulo 4. Enterprise Content Management (ECM) libres

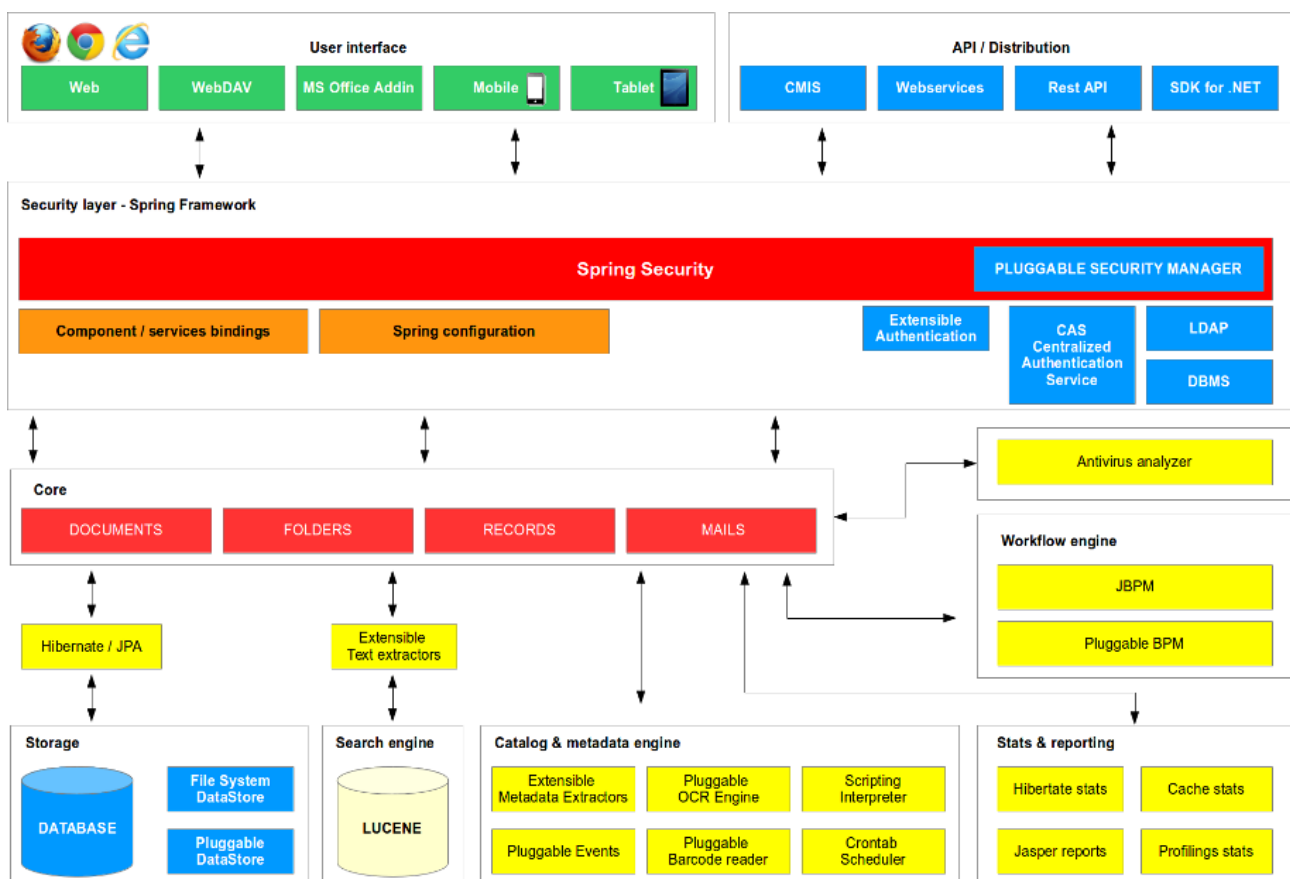
tificaciones a otros usuarios. Una funcionalidad muy interesante es el comparador de versiones, que permite prevvisualizar en la misma pantalla dos versiones diferentes del historial del documento y compararlas de un vistazo. La idea es que el usuario ahorre tiempo en las búsquedas de documentación y pueda acceder a ella en cualquier momento y desde cualquier dispositivo informático, no solamente desde el ordenador de la oficina.

Un aspecto mejorable de OpenKM era la imposibilidad de definir series documentales, lo que hacía que la creación y gestión de las reglas de conservación fuera más compleja que en los demás ECM de código abierto. OpenKM Professional 7 ya ofrecerá a sus usuarios la opción de trabajar con series documentales, si bien esta funcionalidad tardará más tiempo en llegar a la versión Community.

Aunque está optimizado para la gestión documental pura y dura, el programa también permite gestionar contenidos web y se integra de forma nativa con los CMS Wordpress y Joomla. Asimismo, con los conocimientos técnicos adecuados se puede utilizar como back-end para una base de conocimiento, un CRM (Customer Relationship Management) o aplicaciones web desarrolladas a medida.

4.2. Arquitectura

La arquitectura de OpenKM no es muy diferente a la del resto de ECM de código abierto que estamos viendo en este curso. Se trata de un programa desarrollada en Java que se ejecuta como aplicación web en un servidor Tomcat. Esto asegura la compatibilidad con los sistemas operativo más comunes (OpenKM funciona en GNU/Linux, Microsoft Windows, macOS, FreeBSD y UNIX) y evita tanto al usuario como al administrador tener que instalar un cliente pesado en su equipo. El acceso a la interfaz de usuario de OpenKM se realiza a través de un navegador web. Una capa de seguridad media entre la aplicación y el usuario para verificar la identidad de este último.



Arquitectura de OpenKM.
Fuente: *OpenKM – Gestión documental*.

Módulo 4. Enterprise Content Management (ECM) libres

El núcleo del programa está conectado con la base de datos (OpenKM es compatible con HSQL, MySQL, PostgreSQL, MS SQL Server, Oracle, DB2), con el motor de búsqueda Lucene, el antivirus ClamAV, el motor de flujos de trabajo (que puede ser jBPM o Bonita BPM), el OCR (OpenKM es compatible con Tesseract, Cuneiform, ABBYY, Ko-fax, etc.) y otras extensiones. El conjunto de la herramienta se puede integrar con otras aplicaciones y sistemas mediante API, SDK de OpenKM para PHP, .NET y Java, autenticación plugable (PAM) y CMIS.

4.3. Distribución

OpenKM se distribuye en tres versiones: una versión Community, que es totalmente libre y se distribuye bajo licencia GPLv2, y las versiones Cloud y Professional, que incorporan diferentes servicios de pago, entre ellos acceso anticipado a algunas funcionalidades. Una comparativa simplificada de las diferentes versiones puede verse en la siguiente tabla.

	OpenKM Community	OpenKM Cloud	OpenKM Professional
OpenKM	Incluido	Incluido	Incluido
Funcionalidades adicionales	-	Incluido	Incluido
Instalación en la nube	-	Incluido	-
Instalación en servidor de la organización	-	-	Incluido
Configuración	-	Incluido	Incluido
Personalización de la aplicación con el logotipo de la organización	-	Incluido	Incluido
Hosting	-	Incluido	-
Usuarios	Ilimitados	Desde 5 usuarios	Desde 10 usuarios
Soporte de la comunidad	Incluido	Incluido	Incluido
Soporte profesional	-	Incluido	Incluido
Documentación	Incluido	Incluido	Incluido
Formación nivel usuario	-	Incluido	Incluido

4.4. Debilidades

Las principales debilidades de OpenKM son las siguientes:

1. Algunas funcionalidades llegan antes a las versiones de pago que a la versión Community.
2. La opción de definir series documentales de momento sólo está disponible en la versión Professional, lo que hace que la creación y gestión de las reglas de conservación es más compleja que en otros ECM de código abierto si nos decantamos por la versión Community o la versión Cloud.
3. No permite dar privilegios de administración a un usuario únicamente sobre un grupo de usuarios.

4.5. Fortalezas

Las principales fortalezas de OpenKM son las siguientes:

1. Es fácil de instalar y empezar a utilizar.

2. Los servicios de pago tienen precios más competitivos que los de Alfresco y Nuxeo.
3. Está bien optimizado.
4. Existe mucha documentación sobre el programa (tanto por escrito como en vídeo) y su comunidad es bastante activa.

5. OpenProdoc

5.1. Descripción

OpenProdoc es un software de gestión documental de código abierto desarrollado por Joaquín Hierro, que buscaba crear una herramienta sencilla, optimizada, flexible y totalmente libre, sin una edición gratuita con funcionalidades recortadas y una edición de pago completa.

La versión 2.0 cuenta con todas las funcionalidades básicas de un software de gestión documental (clasificación y búsqueda de la documentación, control de versiones, gestión de metadatos y reglas de acceso, programación de tareas, etc.) y también permite importar documentos digitalizados con los OCR Kofax Capture o ABBYY FineReader. Además, incorpora una interfaz web totalmente renovada y optimizada y una nueva función de sincronización del repositorio con una carpeta del disco local. A partir de la versión 2.1 —que estará disponible en el último trimestre de 2017—, incorporará también un OPAC (Online Public Access Catalog) para permitir realizar búsquedas de documentos sin tener que autenticarse en la herramienta.

Mientras que otras herramientas requieren conocimientos técnicos para poder parametrizarlas, el panel de administrador de OpenProdoc se presenta en una interfaz muy intuitiva y permite configurar directamente todos los elementos del modelo documental. Además, en la versión 2.0 la administración es totalmente web, por lo que puede realizarse de forma remota sin tener que instalar ninguna herramienta adicional.

Para el usuario final, el uso del software es similar a los exploradores de archivos a los que está acostumbrado: la columna de la izquierda muestra el árbol de carpetas y la columna de la derecha informa sobre el contenido de estas carpetas.

5.2. Arquitectura

OpenProdoc se puede utilizar en cuatro modalidades: portable, cliente Swing, cliente web y, desde la versión 2.0, portable web. Las modalidades portables son la seña de identidad de OpenProdoc: no requieren instalación y pueden ejecutarse en cualquier equipo que disponga de al menos 1 GB de RAM. Esto hace de OpenProdoc un programa muy optimizado y que se puede adaptar a casi cualquier infraestructura informática —desde las más complejas hasta las más rudimentarias—, por lo que podríamos definirlo como un “software documental de guerrilla”.

El software está desarrollado en Java, por lo que puede ejecutarse en cualquier sistema operativo que disponga de una versión de Java 1.6 o superior y funcionar en un servidor de aplicaciones J2EE.

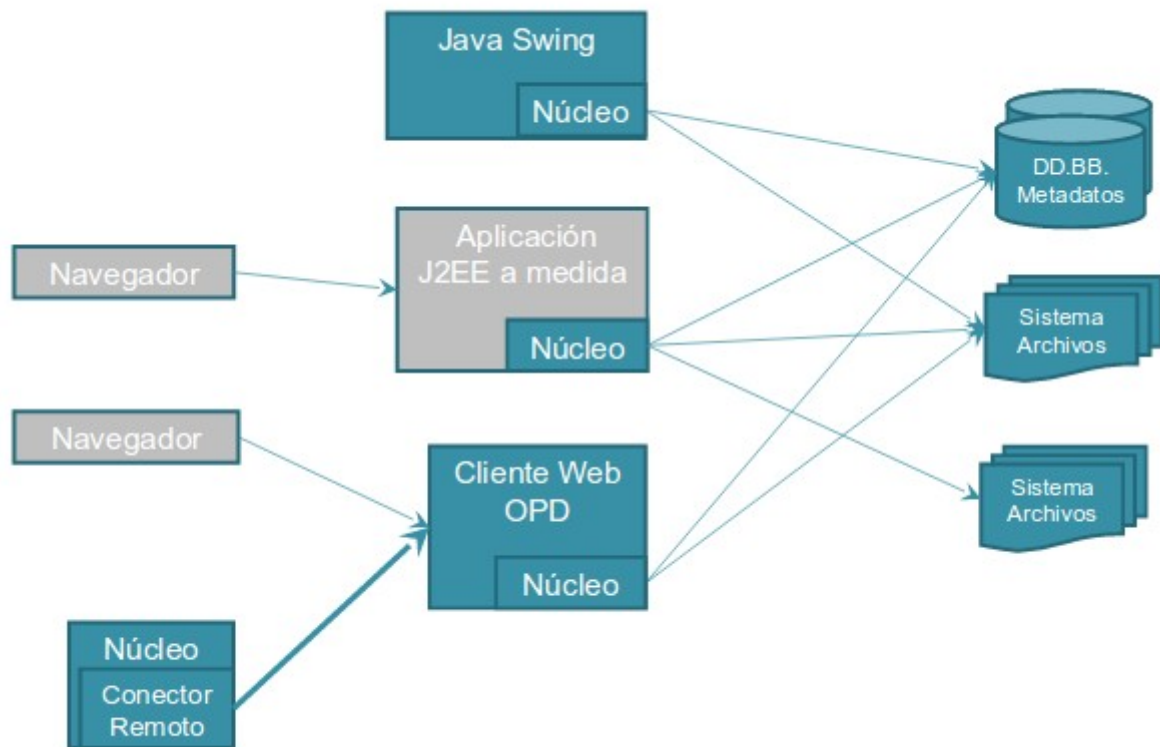
Independientemente de la modalidad, el núcleo de OpenProdoc se compone siempre de una serie de conectores mediante los cuales accede a otros sistemas:

1. Conector de metadatos, que gestiona los metadatos de los documentos.
2. Conector de almacenamiento, que gestiona el almacenamiento y la recuperación de los documentos.
3. Conector autenticador, que valida a los usuarios durante el inicio de sesión.

Para utilizar la modalidad portable, únicamente hay que descomprimir el fichero de OpenProdoc en la carpeta del disco duro o de la memoria externa donde queramos utilizarlo. Al traer embebido el servidor de base de datos HSQLDB, no es necesario instalar tampoco ningún servidor. Esta modalidad permite tener varios usuarios, pero no pueden estar conectados simultáneamente.

Módulo 4. Enterprise Content Management (ECM) libres

No ocurre lo mismo en la modalidad portable web, donde varios usuarios pueden estar conectados a la vez. Para ello, además de la base de datos HSQLDB incluye un servidor J2EE. En esta versión, los usuarios acceden a la interfaz a través de un navegador como Firefox, Chrome o Internet Explorer.



Arquitectura de OpenProdor.

Fuente: Joaquín Hierro (2017).

Las versiones instalables de OpenProdor requieren una base de datos, en la que se almacenará toda la información de la configuración y los documentos. En el caso del cliente web, además hay que configurar un servidor de aplicaciones J2EE. OpenProdor admite como mínimo las bases de datos y servidores de aplicaciones siguientes:

1. Bases de datos:
 1. Derby 10.6, 10.10
 2. MySQL 5.5
 3. PostgreSQL 9.1, 9.4
 4. Oracle 10g
 5. DB2 9.1
 6. MS SQL Server 9.0
 7. HSQLDB 2.2.8
 8. MariaDB 10.0.17
2. Servidores de Aplicaciones:

Módulo 4. Enterprise Content Management (ECM) libres

1. Glassfish 3.1, 4.1
2. Tomcat 6, 7, 8
3. WebLogic Server 11g R1 (10.3.5)
4. JBoss/WildFly 9.0

5.3. Distribución

Las cuatro modalidades de OpenProdoc (portable, portable web, cliente Swing y cliente web) son totalmente libres. Uno de los principios de este proyecto es que absolutamente todo el código sea abierto y, por lo tanto, no existen versiones de pago con funcionalidades recortadas de la versión gratuita.

Actualmente, todo el desarrollo corre a cargo de una única persona, con la que es posible contactar a través de la web del proyecto para solicitar información o soporte. La herramienta continúa recibiendo actualizaciones, la última de ellas en marzo de 2017, y se continúa trabajando en nuevas funcionalidades.

OpenProdoc: creando un gestor documental en solitario (Hierro, 2014)

[He desarrollado OpenProdoc] principalmente como reto profesional y también para tener el Gestor Documental que a mí me gustaría como usuario. He utilizado y desarrollado proyectos sobre muchos gestores y cada uno tiene sus lados fuertes y sus fallos. Se trataba de unir las funciones preferidas de cada uno (las cuales tendrán distinta valoración según cada persona). Adicionalmente, cuando comencé no había muchos gestores Open Source, y me parecía necesario cubrir esa carencia.

Por último, como forma de aprender, ya que no había desarrollado un gestor documental desde cero, y eso implica analizar aspectos de bajo nivel y diseño que no suelen tenerse en cuenta habitualmente pues ya vienen dados en los proyectos por el producto sobre el que se trabaja.

5.4. Debilidades

Las principales debilidades de OpenProdoc son las siguientes:

1. Actualmente no permite ejecutar flujos de trabajo ni se integra con motores de BPM.
2. No se integra con OCR open source.
3. Aunque este programa ha comenzado a recibir soporte comercial por parte de dos empresas, el número de servicios asociados a este producto y de proveedores que los ofrecen es más limitado que en los otros ECM.

5.5. Fortalezas

Las principales fortalezas de OpenProdoc son las siguientes:

1. Absolutamente todo el código del programa es libre: no existe una versión de pago con funcionalidades recortadas de la versión gratuita.
2. Estamos ante un programa muy optimizado, flexible y escalable.
3. OpenProdoc es un ECM sencillo, con una curva de aprendizaje rápida tanto para el usuario como para el administrador. No requiere conocimientos de programación para diseñar el modelo documental y automatizar tareas básicas como la eliminación de documentos.
4. Ofrece una mayor granularidad que otros programas a la hora de asignar permisos a los usuarios.

Resumen

Los ECM son programas que gestionan electrónicamente todas las evidencias generadas por una organización en el desarrollo de su actividad, incluidos contenidos que quedan fuera de la representación tradicional del documento, como los correos electrónicos, los registros o los contenidos web.

La ventaja de las alternativas libres y de código abierto con respecto a los ECM privativos son:

1. Menor coste total.
2. Arquitectura basada en estándares abiertos.
3. Independencia de la organización con respecto al proveedor y la propia herramienta.

Hemos visto que las funcionalidades centrales de los ECM open source son muy similares y que, desde el punto de vista de la arquitectura, todas ellas están desarrolladas en Java, se ejecutan en servidores y contenedores de aplicaciones como OSGi, Tomcat, JBoss y Jetty e interactúan con un amplio ecosistema de sistemas y aplicaciones que las convierten en plataformas modulares e integrables según las necesidades de la organización usuaria.

Donde más se diferencian entre ellas estas herramientas es en las funcionalidades adicionales, los servicios que ofrecen y el grado de compromiso con el software libre y de código abierto. Este último aspecto es fundamental, ya que determina la calidad de la documentación técnica; si existe una versión del software con funcionalidades recortadas de la versión libre o de acceso anticipado para suscriptores de pago o, por el contrario, todo el código del programa es libre; así como el nivel de participación de la comunidad en el desarrollo del programa.

Bibliografía

- Alfresco One 5.1.3 Developer guide [web]. [Consulta 05-05-2017]. Disponible en: <http://docs.alfresco.com/5.1/concepts/dev-for-developers.html>.
- BYRD, M. "90 days in – Attend new ECD Product Roadmap Webinar". En: *OpenText* [blog]. 19 abril 2017 [consulta 05-05-2017]. Disponible en: <https://community.emc.com/docs/DOC-57638>.
- CONRADT, M. 2017. "Dumping Alfresco for Nuxeo as my preferred Document Management System". En: *Medium* [en línea]. 24 enero 2017 [consulta 05-05-2017]. Disponible en: <https://medium.com/mathias-conradt/dumping-alfresco-for-nuxeo-as-my-preferred-document-management-system-2116cf421d36>.
- HIERRO, J. 2014. "OpenProdoc: creando un Gestor Documental en solitario". En: *Dokumentalistas* [en línea]. 26 febrero 2014 [consulta 05-05-2017]. Disponible en: <http://www.dokumentalistas.com/articulos/openprodoc-creando-un-gestor-documental-en-solitario/>.
- *Gestión documental, tecnologías ECM y OpenProdoc* [en línea]. 19 abril 2017 [consulta 05-05-2017]. Disponible en: <https://jhierrot.github.io/openprodoc/Docs/PresentacionOpenprodoc2.0.pdf>.
- HUMBLE, C. "Catching up with Nuxeo: Switching from Python to Java". En: *InfoQ* [en línea]. 9 agosto 2010 [consulta 05-05-2017]. Disponible en: https://www.infoq.com/articles/nuxeo_python_to_java/.
- Nuxeo Documentation Center Home [en línea]. [Consulta 05-05-2017]. Disponible en: <https://doc.nuxeo.com/nxdoc/quick-overview/>.
- OpenKM – *Gestión documental* [en línea]. [Consulta 5 de mayo 2017]. Disponible en: <https://www.openkm.com/es/arquitectura.html>.
- POTTS, J. "What's going on with Alfresco clustering?" En: *ECM Architect* [blog]. 17 octubre 2012 [consulta 05-05-2017]. Disponible en: <http://ecmarchitect.com/archives/2012/10/17/1745>.
- POTTS, J. "The plain truth about Alfresco's open source ethos". En: *ECM Architect* [blog]. 24 mayo 2015 [consulta 05-05-2017]. Disponible en: <http://ecmarchitect.com/archives/2015/05/24/4017>.
- UNE-ISO 15489:2016. Información y documentación. Gestión de documentos. Parte 1: Conceptos y principios.
- UNE-ISO 30300:2011. Información y documentación. Sistemas de gestión para los documentos. Fundamentos y vocabulario.