

TEMA 104. GESTIÓN DOCUMENTAL. GESTIÓN DE CONTENIDOS. TECNOLOGÍAS CMS Y DMS DE ALTA IMPLANTACIÓN

Actualizado a 12/05/2023.

RESUMEN

1	TABLA DE CONTENIDO	
1	Tabla de contenido	2
2	Gestión Documental	2
2.1	Conceptos y Evolución	2
2.2	Ejemplos de Gestores documentales	4
2.3	Arquitecturas Funcionales	6
3	Gestión de Contenidos	8
3.1	Conceptos y Evolución	9
3.2	Arquitecturas Funcionales	11
3.3	Ejemplos de Gestores de contenidos	12
4	Estándares y normas	13
5	Sindicación	15
6	Referencias	16

2 GESTIÓN DOCUMENTAL

La gestión documental o gestión de documentos, es el conjunto de normas técnicas y prácticas usadas para **administrar los documentos de todo tipo**, recibidos y creados en una organización, facilitar la recuperación de información desde ellos, determinar el tiempo que los documentos deben guardarse, eliminar los que ya no sirven y asegurar la conservación a largo plazo de los documentos más valiosos, aplicando principios de racionalización y economía.

Por tanto, los Gestores Documentales (SGD Sistemas de Gestión Documental o DMS Document Management Systems) son herramientas que permiten administrar el flujo de documentos (o activos digitales) de todo tipo en una organización realizando todas las tareas arriba indicadas. En definitiva, facilitan y gestionan el almacenamiento seguro, la indexación y la recuperación de contenidos documentales. Están especializados en el escaneado de documentos, indexación, optimización del almacenamiento, archivo y gestión de procesos y flujos de trabajo.

Se pueden **clasificar** por su **lenguaje de programación** (Java etc) o por el **tipo de licencia** (código abierto vs propietario)

2.1 CONCEPTOS Y EVOLUCIÓN

Los SGD han evolucionado bastante desde su creación, cambiando la terminología, alcance y tecnología en la que se basan.

Gestión de Registros (Record Management): Los documentos del tipo record son los que contienen información que se debe conservar para presentar ante inspecciones, auditorías o los tribunales.

Definición del documento, tipo record, según la ISO/UNE 15489: «Información creada o recibida, conservada como información y prueba, por una organización o un individuo en el desarrollo de sus actividades, o en virtud de sus obligaciones legales». Ejemplos: facturas, contratos, presupuestos, planos de la empresa, vídeos y fotografías de presentaciones de productos, logos, páginas web, bases de datos etc.

Gestión de Contenido Empresarial (ECM Enterprise Content Management): concepto que engloba no solo la gestión de documentos sino también la estrategia empresarial, métodos y herramientas para gestionar los documentos empresariales relacionados con los procesos de negocio. Estos sistemas incluyen la arquitectura tecnológica que permite su explotación.

Los ECM comenzaron su implantación masiva a partir del año 2000, definiendo metodologías y mejores prácticas para la captura, análisis, mapeo, conservación y almacenamiento de información. Aparecieron los sistemas colaborativos, nuevos canales de acceso a la información, y las organizaciones perciben que tienen numerosos silos de información que complican la gestión y compartición real de los documentos y su información (datos).

Los ECM tienen que evolucionar para dar respuesta a los nuevos requisitos de negocio, de los clientes, y de la tecnología para dar respuesta a la multicanalidad (omnichannel). ECM ha pasado a ser una especie de paraguas o esqueleto que permite implementar la estrategia de contenidos (digitales y físicos) de una organización. Incluso Gartner declaró el término ECM como obsoleto en 2017 para evolucionar hacia los Content Services (Servicios de Contenidos), con una tecnología moderna basada en:

- Basado en Cloud: Contenedores, microservicios
- Bases de datos NoSQL y SQL
- Open Source
- API REST
- Pocas necesidades de desarrollo de código

En paralelo, surgen necesidades de gestión de contenidos digitales, creados y mantenidos para los canales digitales, que tienen otras necesidades de gestión como se verá en la segunda parte del tema. (SGC Sistemas de Gestión de Contenidos o CMS Content Management Systems).

2.2 EJEMPLOS DE GESTORES DOCUMENTALES

Herramienta	Características		
Alfresco (también se considera gestor de contenidos)	<ul style="list-style-type: none"> Tiene integrado el motor jBPM (definición de procesos) y el motor Activiti Está basado en la JSR 170 (gestión documental) Dispone de soporte comercial en su versión Enterprise Proporciona acceso al repositorio mediante CIFS, FTP y WebDAV Posibilidad de instalación en clúster Altamente escalable Especialmente pensado para integrarse con Liferay Portal 	Java	Código abierto
Magnolia	<ul style="list-style-type: none"> Multiplataforma Proporciona módulos que integran frameworks como struts, spring, stripes o wickets para facilitar la integración con aplicaciones propias del usuario Integra JAAS (estándar Java para la autenticación y autorización de usuarios) 	Java	Código abierto
Documentum	<ul style="list-style-type: none"> Multiplataforma Proporciona acceso al repositorio mediante servicios web, WebDAV, FTP y SMB/CIFS 	Java	SW propietario
OpenKM		Java	Código abierto
OpenText		Java	SW propietario
MS Sharepoint		C#	SW propietario

Otros gestores documentales son: DocuWare, Docufree, Dokuflex, LogicalDOC, Laserfiche, SARIP, Skysite, Nuxeo, R2 Docuo, GenieBelt, PaperTracer, Process Street, Zoho Docs.

Entre las alternativas reutilizables de la Administración relacionadas con la gestión documental están:

- **Espacios de Colaboración** (Tecnología: Java. Basado en Liferay. Responsable: MINHAP): Presenta funcionalidades como gestión documental, foros y agenda..
- **Inside** (Tecnología: Java. Responsable: MINHAP): permite almacenar y modificar documentos y expedientes electrónicos en cualquier gestor documental que sea compatible con el estándar CMIS, así como los metadatos obligatorios asociados. (Ver documentación Adicional)
- **G-Inside** (Tecnología: SaaS con web services. Responsable: MINHAP): permite la generación de documentos electrónicos y expedientes electrónicos conformes con el ENI.
- **@Doc** (Tecnología: Java. Responsable: Ministerio de la Presidencia): plataforma de documento y expediente electrónico.

También **"ARCHIVE"**, aplicación web de archivo definitivo (Ver documentación Adicional)

Productos de gestión documental

- Alfresco
- Documentum
- DocuWare
- Docufree
- DocuShare
- Dokuflex
- GenieBelt
- Laserfiche
- LogicalDOC
- OnBase Cloud
- OpenKM
- PaperTracer
- Process Street
- R2 Docuo
- SARIP
- SKYSITE
- Zoho Docs
- Nuxeo

Cuadrante mágico de Gartner 2020 para plataformas de Content Services (evolución de ECMs)

Figure 1: Magic Quadrant for Content Services Platforms



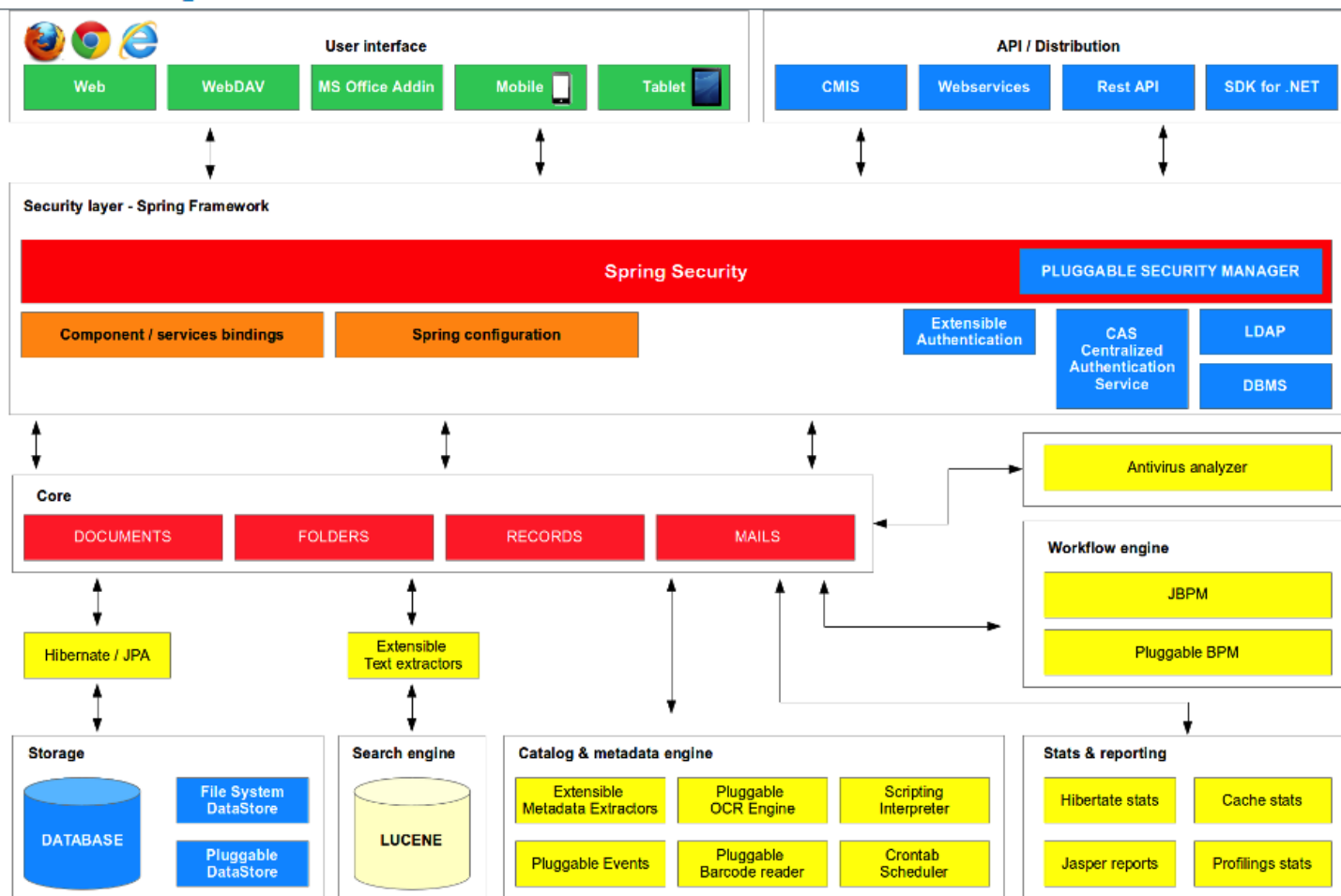
Source: Gartner (November 2020)

2.3 ARQUITECTURAS FUNCIONALES

Con el objeto de facilitar la comprensión de estos sistemas, se muestra un ejemplo tipo de arquitectura funcional de un producto de mercado, considerando que la aproximación arquitectural es muy similar en todos ellos, y que la gran mayoría cumplen con los estándares y normas que se indican al final de este documento.

La siguiente imagen muestra la [arquitectura de OpenKM](#) y sus diferentes capas.

Diagrama de la arquitectura del sistema



En esta otra imagen se muestra la arquitectura de Alfresco Content Platform

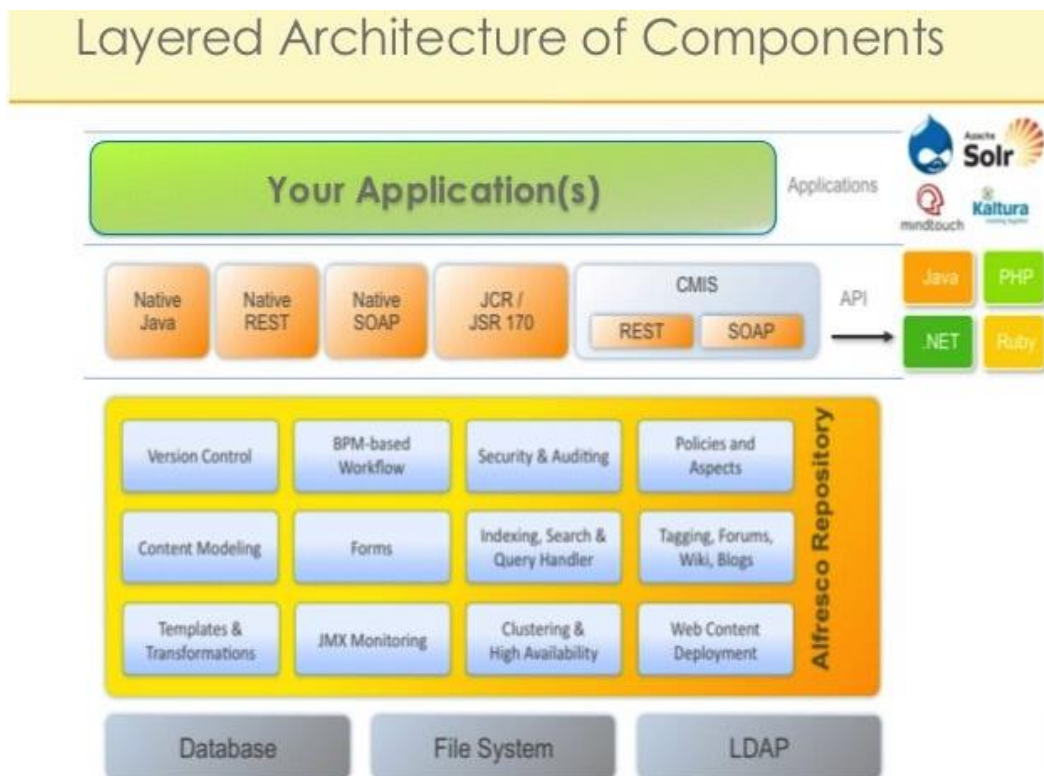


Figura 1: arquitectura de Alfresco Content Platform

3 GESTIÓN DE CONTENIDOS

Un sistema de **gestión de contenidos** o **CMS (content management system)** es una aplicación sw con un entorno de trabajo para la creación y publicación, gestión y administración de contenidos accesibles digitalmente (páginas web y desde diferentes dispositivos), por parte de los administradores, editores, participantes y demás usuarios.

Los CMS permiten crear usuarios con diferentes perfiles y diferentes niveles de acceso (p.e. revisores, publicadores de contenido, etc) para simplificar y controlar los estados por los que pasa cada contenido desde su edición, su revisión, hasta su publicación. **Facilita** por tanto **la gestión de los contenidos y simplifica su administración**.

Un ejemplo clásico es el de editores que cargan el contenido al sistema y otro de nivel superior (moderador o administrador) que permite que estos contenidos sean visibles a todo el público (los aprueba).

Cuenta con una **interfaz** que controla una o varias **bases de datos donde se aloja el contenido**. El sistema **permite manejar de manera independiente el contenido y el diseño**. Así, es posible manejar el contenido y darle en cualquier momento un diseño distinto (sin tener que modificar el contenido cada vez)

Todo Sistema Gestor de Contenidos realiza el siguiente proceso de gestión:

- **Creación** de la información.- El usuario genera una nueva información, que podrá ser de diverso tipo (texto, gráficos, imágenes, etc.)
- **Presentación** de la información
- **Publicación** de la información
- **Mantenimiento** de la información: actualización dentro de la cual se puede editar o borrar un determinado contenido.

Se pueden **clasificar** por su **lenguaje de programación** (Java, Ruby on Rails, Python etc) o por el **tipo de licencia** (código abierto vs propietario)

Hay algunos productos que pueden ser empleados como CMS pero que debido a sus funcionalidades más amplias se pueden encuadrar como gestores de portales – que permiten la creación y administración de portales web – como por ejemplo: Liferay, Microsoft Office SharePoint Server o Wordpress.

Pueden tener ciertas características como;

- Soporte para multi idioma
- Generación de estadísticas y reporting
- Gran cantidad de módulos: foros, blogs, encuestas,RRSS...
- Control de versiones
- Plantillas predefinidas
- Sistema de permisos basado en roles/grupos
- Sistemas de cacheo
- Urls amigables
- Motor de búsqueda
- Compatibilidad con SSO (Single sign on)
- ...

Los Gestores de Contenidos pueden ser interoperables si cumplen con el **estándar CMIS (Content Management Interoperability Services** o Servicios de Interoperabilidad de Gestión de Contenidos).

Los CMS también se conocen con el nombre de Web CMS. Las versiones modernas se integran con funcionalidades analíticas y de Big Data para obtener mayor y mejor información de los usuarios de los sistemas con vistas a la aplicación de campañas de marketing, personalización de contenidos, mejora de la experiencia de usuario, ...

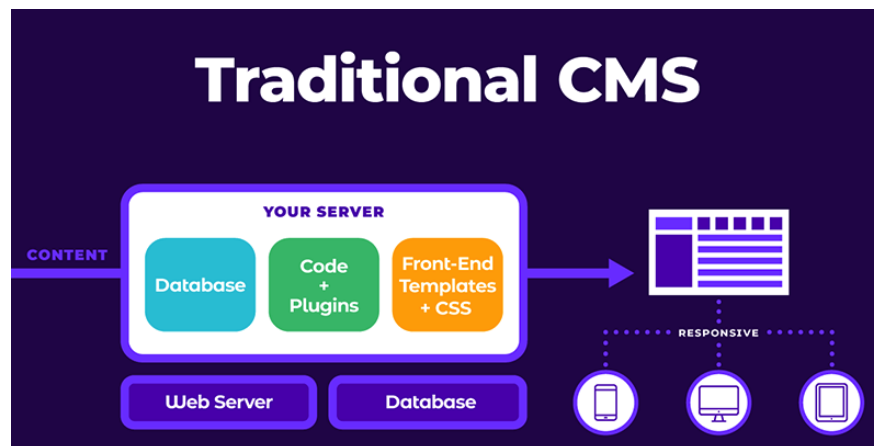
3.1 CONCEPTOS Y EVOLUCIÓN

Los CMS han evolucionado hacia diferentes modelos de arquitectura, tanto desde el punto de vista de su ubicación (en los CPDs de los clientes a servicios en Cloud, tanto públicos como privados).

Actualmente conviven 3 aproximaciones diferentes:

1. Sistemas Tradicionales:

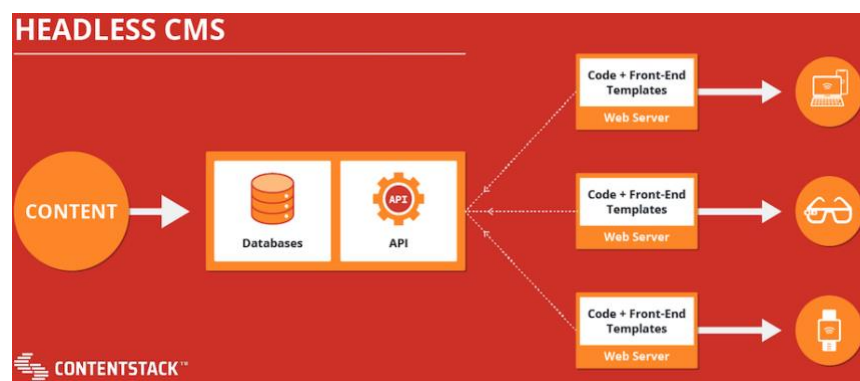
- a. Ofrecen en una misma arquitectura toda la funcionalidad del gestor de contenidos,



con componentes bastante acoplados.

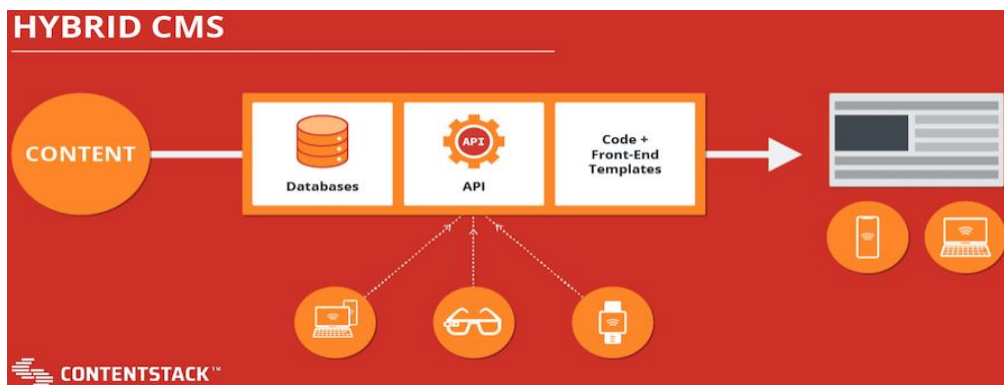
2. Sistemas Headless (sin cabecera o desacoplada)

- a. Estos sistemas tienen una arquitectura que desacopla el frontend del backend, lo que permite almacenar datos en un lugar y enviarlos o distribuirlos a través de múltiples canales y servicios (Multi canalidad).
- b. Se apoya en la arquitectura de microservicios, aportando mayor flexibilidad para adaptarse a los requisitos cambiantes, aunque al mismo tiempo requiere la construcción de plantillas ad-hoc.



3. Sistemas híbridos:

- a. Algunas organizaciones requerirán modelos híbridos, integrando ambos escenarios.
- b. El modelo híbrido incluye una capa de presentación similar al tradicional al tiempo que permite un desacoplamiento para ciertos canales a través de las APIs.



3.2 ARQUITECTURAS FUNCIONALES

Ejemplo de arquitectura funcional de un CMS - Drupal

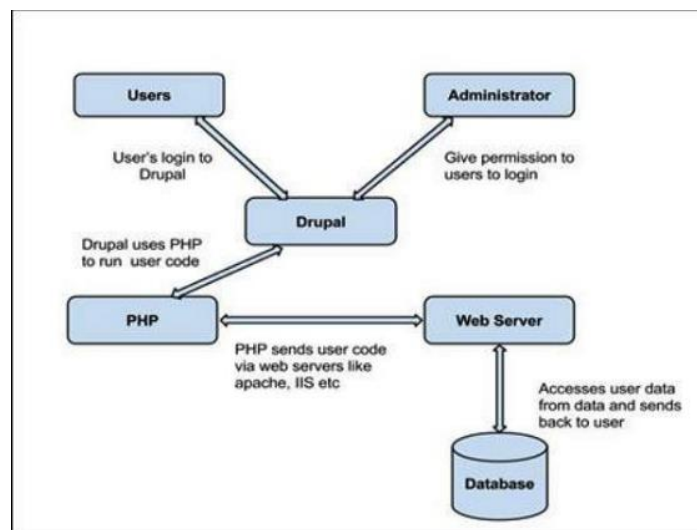


Figura 2: Ejemplo de arquitectura funcional de un CMS - Drupal

Arquitectura Drupal para la creación y publicación de contenidos

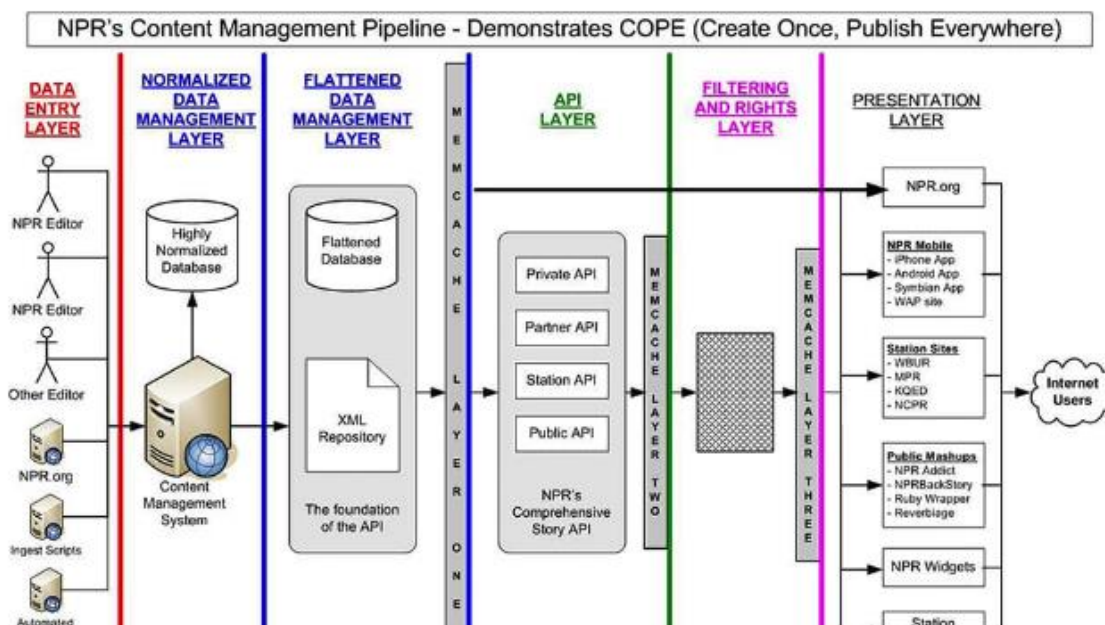


Figura 3: Arquitectura Drupal para la creación y publicación de contenidos

Photo credit: NPR: Create Once Publish Everywhere (Slide #30)

3.3 EJEMPLOS DE GESTORES DE CONTENIDOS

Algunos ejemplos de Gestores de Contenidos son:

Herramienta	Tecnología	Licencia
Drupal	PHP	Código abierto (GPL)
Joomla	PHP	Código abierto (GPL)
OpenCMS	Java	Código abierto (LGPL)
Wordpress	PHP	Código abierto (GPL)
Magnolia	Java	Código abierto
IBM WebContent Management	Java	SW propietario

Otros gestores de contenidos son:

Blogs:	Foros:	Portales:	Wikis:	Educación:	Comercio:	Almacenamiento de archivos:
<ul style="list-style-type: none"> • Movable Type • Textpattern • Wordpress • Blogger • Tumblr 	<ul style="list-style-type: none"> • MyBB • phpBB • punBB • Simple Machines Forum 	<ul style="list-style-type: none"> • Apache Lenya • Kentico CMS • Drupal • Jaws • Joomla! • Mambo • Plone • XWiki • Magnolia CMS 	<ul style="list-style-type: none"> • Dokuwiki • MediaWiki • TiddlyWiki • WikkaWiki 	<ul style="list-style-type: none"> • Mahara • Moodle • Claroline 	<ul style="list-style-type: none"> • osCommerce • Magento • PrestaShop 	<ul style="list-style-type: none"> • ownCloud • Dataprius

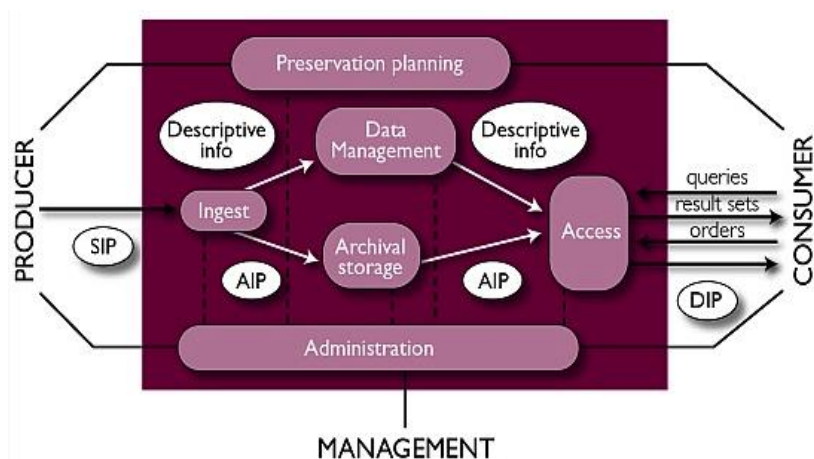
Una solución de la Administración es “**ACCEDA**” (Ver documentación Adicional)

Para más información al respecto, se puede consultar la documentación de la carpeta “Doc Adicional”.

4 ESTÁNDARES Y NORMAS

OAIS (Open Archival Information System): modelo conceptual que define los procesos necesarios para la gestión, archivado y preservación a largo plazo de documentos. Norma ISO 14721.

El [modelo de referencia se OAIS](#) define la ingesta, acceso, administración, gestión de datos, preservación y almacenamiento. La siguiente figura ilustra el proceso.



Los tipos de paquetes de información son:

- **SIP (Submission Information Package):** paquete de información que se envía al sistema para su ingesta, contiene el objeto original transferido por el productor (datos y metadatos).
- **AIP (Archival Information Package):** información de archivo en que se transforma un SIP para su almacenamiento.

- **DIP (Dissemination Information Package):** información de difusión, que contiene el objeto con la respuesta digital a una solicitud.

El servicio compartido ARCHIVE (AGE) de archivo electrónico cumple el estándar OAIS.

ISO 15489 (Gestión de documentos): norma que regula la gestión de documentos que producen las organizaciones. La norma ofrece una metodología para el diseño e implementación de un DMS.

ISO 30300 (Sistema de gestión de documentos): da continuidad a la norma ISO 15489: La norma propone un sistema para la gestión de los documentos basado en la mejora continua.

MoReq (MODELO EUROPEO DE REQUERIMIENTOS PARA UN SISTEMA DE GESTIÓN DE DOCUMENTOS ELECTRÓNICOS DE ARCHIVO): Norma gestión de documentos electrónicos de archivo. Elaborado por DLM-Forum, Unión Europea (programa IDA). Este documento describe con detalle los requisitos de los sistemas de gestión de registros electrónicos que propone esta especificación, sus sistemas de clasificación de estos registros, de control y seguridad, conservación y eliminación, búsqueda, recuperación y eliminación, así como los requisitos no funcionales y los referentes a los metadatos.

Content Repository API for Java (JCR) es una especificación para una interfaz de programación de aplicaciones (API) de la plataforma Java para acceder a los repositorios de contenido de manera uniforme. La especificación fue desarrollada bajo el Java Community Process como **JSR-170** (Versión 1), y como **JSR-283** (versión 2). El paquete Java es javax.jcr. Jackrabbit es la implementación de Apache de JCR. Gestores de contenidos como Magnolia (se usa en la administración) o Liferay lo utilizan.

CIFS (Common Internet File System, Microsoft): protocolo red capa aplicación, proporcionar acceso compartido a ficheros.

WebDAV (Web Distributed Authoring and Versioning): extensión del protocolo HTTP para crear, cambiar y mover documentos en un servidor remoto. Estos documentos se almacenan en unidades virtuales (web folders).

CMIS (Content Management Interoperability Services o Servicios de Interoperabilidad de Gestión de Contenidos) es un estándar abierto que permite a diferentes sistemas de gestión de contenidos interoperar a través de Internet. En concreto, CMIS define una capa de abstracción para el control de diversos sistemas de gestión de documentos y repositorios, que utilicen protocolos web. El estándar CMIS admite operaciones básicas de administración de documentos, como crear, leer, actualizar, eliminar, proteger y desproteger, así como de administración de versiones de documentos y sus metadatos.

CMIS ofrece interoperabilidad entre las API que lo admiten, pero no sustituye a las API nativas.

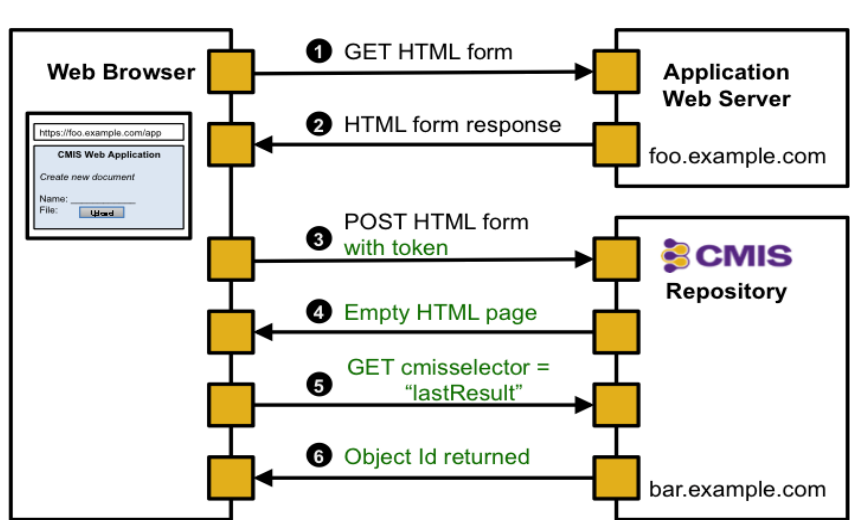


Figura 4: Ejemplo de interacción web basada en CMIS

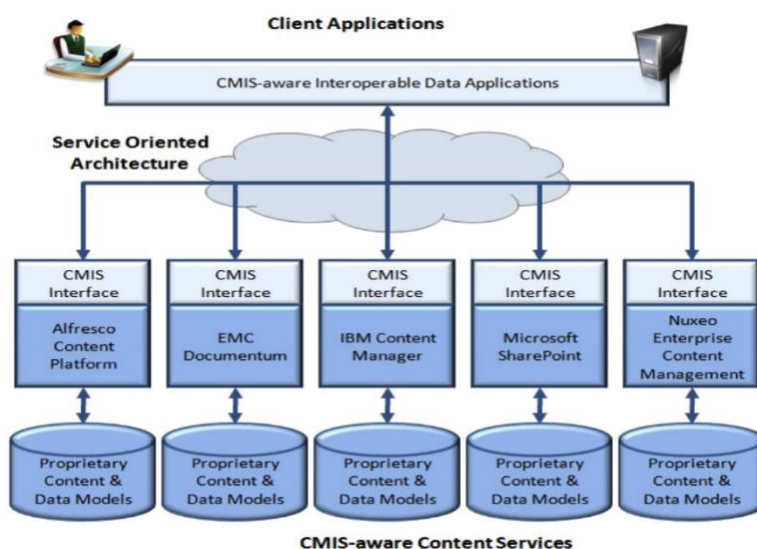


Figura 5: Ejemplo de arquitectura de servicios basada en CMIS

5 SINDICACIÓN

Redifusión web (o sindicación web) es el reenvío o reemisión de contenidos desde una fuente original (sitio web de origen) hasta otro sitio web de destino (receptor) que a su vez se convierte en emisor puesto que pone a disposición de sus usuarios los contenidos a los que en un principio sólo podían tener acceso los usuarios del sitio web de origen.

Las dos principales familias de formatos de redifusión web son RSS y Atom.

RSS son las siglas de **Really Simple Syndication**, un formato XML para distribuir contenido en la web. Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenidos. El formato permite distribuir contenidos sin necesidad de un navegador, utilizando programas llamados agregadores de noticias, diseñados para leer contenidos RSS.

RSS, Esta sigla se usa para referirse a los siguientes estándares:

- Rich Site Summary (RSS 0.91) "Resumen óptimo del sitio"
- RDF Site Summary (RSS 0.9 y 1.0) "Resumen RDF del sitio"
- Really Simple Syndication (RSS 2.0) "Sindicación Realmente Simple"

Algunos agregadores web populares:

- Bloglines
- MyYahoo!
- Netvibes
- FeedReader
- Feedly

El nombre **Atom** hace referencia a dos estándares relacionados.

- Formato de Redifusión Atom es un fichero en formato XML usado para Redifusión web.^[7]
- Protocolo de Publicación Atom (resumido en Inglés AtomPub o APP) es un protocolo simple basado en HTTP para crear o actualizar recursos en Web.^[8]

El formato Atom fue desarrollado como una alternativa a RSS y publicado como un "estándar propuesto" de la IETF en el RFC 4287, y el protocolo de comunicación se publicó como RFC 5023.

6 REFERENCIAS

- [PAE] Herramientas para la gestión de documentos y expedientes electrónicos:
https://administracionelectronica.gob.es/pae_Home/pae_Estrategias/Archivo_electronico/pae_herramientas_para_la_gestion_de-documentos_y_expedientes_electronicos0.html
- [PAE] INSIDE:
https://administracionelectronica.gob.es/ctt/inside#.X2eW_JMzY1I
- [PAE] ARCHIVE:
<https://administracionelectronica.gob.es/ctt/archive#.X2eXAJMzY1I>
- [PAE] ACCEDA:
<https://administracionelectronica.gob.es/ctt/acceda#.X2eXaZMzY1I>