



**Consejería de Innovación, Ciencia y Empresa  
Servicio de Innovación y Prospectiva**

**Plataforma de Interoperabilidad de Contenidos Web (iCMS)  
Comunicación Tecnimap 2010**

Autores:

Juan Leal Zubiete

Carlos Leal Zubiete

Gustavo Vicente Sánchez

Francisco Rodríguez Corredor

Pedro León Mora

Gregorio Mora Hidalgo

Antonio Santos Velasco

Empresa desarrolladora: Viavansi S.A

## ÍNDICE

1	Temario de Comunicaciones.....	3
1.1	Implicaciones económicas y mercado único.....	3
1.2	Eficiencia y sostenibilidad.....	3
1.3	Iniciativas legales y tecnológicas.....	3
2	Comunicación.....	4
2.1	Introducción.....	4
2.2	La Interfaz iCMS.....	5
2.2.1	Niveles y Estándares.....	5
2.2.2	Arquitectura.....	6
2.3	Funcionalidad.....	7
2.4	Rendimiento.....	7
2.4.1	Configuración Hardware.....	8
2.4.2	Configuración Software.....	8
2.4.3	Prueba 1.....	8
2.4.4	Prueba 2.....	8
2.4.5	Valoración final.....	9
2.5	Próximos hitos.....	9

## 1 Temario de Comunicaciones

El presente trabajo se engloba dentro de las siguientes áreas de interés incluidas en el congreso:

### 1.1 Implicaciones económicas y mercado único

- **Servicios interoperables:** este producto está enfocado principalmente a la interoperabilidad entre distintos sistemas ya sean fuentes de datos o sistemas de representación de los mismos, pero no sólo como arquitectura orientada a servicios (SOA) sino también desde el punto de vista de la semántica.

### 1.2 Eficiencia y sostenibilidad

- **Interoperabilidad entre Administraciones Públicas:** la interfaz desarrollada en este proyecto nos permitirá la publicación de contenidos y posterior explotación entre distintos sistemas de información. Dicha comunicación se puede realizar tanto de forma interna, a través de la Plataforma de Interoperabilidad de la Junta de Andalucía sobre la red Nerea, como externa, es decir, entre distintas administraciones públicas.
- **Reutilización de información y servicios:** al basarse en tecnologías y protocolos relacionados con la web semántica (RDF, SPARQL, OWL) y en los principios de diseño de *linked data*, la interfaz iCMS facilita la reutilización de los datos presentes en sistemas de información heterogéneos. De esa forma, se promueve el desarrollo de nuevos servicios y visualizaciones que combinen fuentes de datos diversas, tanto propias de la organización como externas.

### 1.3 Iniciativas legales y tecnológicas

- **Aplicaciones de innovaciones tecnológicas para las Administraciones Públicas:** la utilización de este proyecto por parte de la Junta de Andalucía ya se puede considerar en sí una innovación tecnológica en el seno de la Administración Pública, debido al uso de tecnologías de web semántica. Además de ello se propone un enfoque global al problema de la interconexión entre sistemas de información relacionados con la publicación web.

## 2 Comunicación

### 2.1 Introducción

En la Junta de Andalucía existe un número muy elevado de aplicaciones que gestionan información susceptible de publicarse con tecnologías web, ya sea en portales públicos o en intranets privadas.

Actualmente, esta información reside en aplicaciones concebidas para diversos fines: gestores de contenidos, gestores documentales, sistemas informáticos de apoyo a la tramitación administrativa, bases de datos corporativas... Los entornos tecnológicos en los que están desarrolladas estas aplicaciones, las operaciones que permiten y los formatos de almacenamiento son también variados.

Existen también diversas aplicaciones corporativas que necesitan acceder a esta información para prestar servicios a los usuarios de manera horizontal: gestores de presentación web, gestores de contenidos, motores de búsqueda... Esta necesidad resulta evidente en el caso del portal institucional de la Junta de Andalucía, cuya misión es ofrecer un acceso común a la información y servicios ofrecidos por la administración andaluza a través de Internet de la forma más útil y sencilla para la ciudadanía.

La tarea de integrar distintos sistemas de información se ha afrontado tradicionalmente en la Junta de Andalucía mediante el desarrollo de conectores *ad hoc*. No obstante, este enfoque resulta problemático desde el punto de vista de la eficiencia, ya que implica realizar un desarrollo informático específico para cada grupo de sistemas que deban interactuar conjuntamente, trasladando además a cada caso particular cuestiones como el rendimiento, capacidad, seguridad o escalabilidad de la solución adoptada. Frente a ello, la Junta de Andalucía propone un enfoque global al problema de la interconexión entre sistemas de información relacionados con la publicación web.

Este tipo de planteamientos no es exclusivo de los contenidos web: de hecho, la solución propuesta se enmarca en el proyecto de Plataforma de Interoperabilidad de la Junta de Andalucía, que pretende servir como infraestructura de soporte común para facilitar la integración e interoperabilidad entre servicios de la Junta de Andalucía y de las distintas Administraciones, a partir de una arquitectura orientada a servicios (SOA).

El interfaz de interoperabilidad para contenidos web iCMS prescribe un conjunto de protocolos, operaciones y formatos abiertos para el intercambio de contenidos web entre diversos sistemas internos de la Junta de Andalucía. Al optar por una solución común, se limita el número de implementaciones necesarias para interconectar entre sí todas las fuentes de información y consumidores relevantes reutilizando desarrollos. Al emplear estándares abiertos, facilita su extensión a otros escenarios relacionados con la interoperabilidad de contenidos web no previstos inicialmente.

Además del interfaz, se ha desarrollado una implementación de referencia definida como una librería Java, con el objetivo de facilitar su integración en los sistemas corporativos en los que resulte necesario. Esta implementación se distribuye como software libre a través del Repositorio de Software de la Junta de Andalucía.

## 2.2 La Interfaz iCMS

La Interfaz de Interoperabilidad iCMS consigue la interconexión de sistemas que en la actualidad utiliza la Junta de Andalucía haciendo uso de diversos protocolos estándares. Su principal función es el consumo y gestión de información de un sistema por parte de otro utilizando un marco de comunicación estándar.

Para ello, se ha optado por emplear el enfoque de la web semántica como el camino para conseguir la interoperabilidad de contenidos entre diferentes sistemas. Así, los contenidos que se manejan y transmiten entre los diferentes sistemas iCMS vienen descritos por metadatos, que les aportan un significado más complejo y permiten operaciones potentes (como una búsqueda más inteligente) sobre ellos.

La Interfaz de Interoperabilidad establece un canal de comunicación al definir un medio de acceso a datos común para todos los sistemas, un lenguaje de descripción de elementos constante y un conjunto de operaciones que permiten la interacción funcional entre los sistemas. Toda la comunicación se hace a través del protocolo HTTP, permitiendo el acceso a través de una interfaz de tipo REST o como servicio web.

### 2.2.1 Niveles y Estándares

En el proyecto se han definido tres formas de interoperabilidad, llamadas niveles, destinadas a cubrir escenarios de uso distintos:

- Nivel A: ligero, escalable, con rendimiento elevado, sencillo de implementar y muy alineado con los estándares de web semántica, aunque con un espectro limitado de operaciones (ej. no soporta operaciones de escritura). Se ha diseñado con el objeto de que se puedan realizar implementaciones eficientes en alta carga y que proporcionen un rendimiento excelente.
- Nivel B: La interfaz de Nivel B está pensada para garantizar la integración con cualquier sistema de información según las pautas definidas en la Plataforma de Interoperabilidad de la Junta de Andalucía. Añade asimismo la funcionalidad de publicación.
- Nivel C: El nivel C es una extensión del nivel A en la cual las consultas que se pueden realizar son más complejas, ya que se permite al cliente el uso del lenguaje SPARQL en la petición. En la implementación desarrollada aún no se contempla este nivel.

Para la comunicación entre sistemas de la Interfaz, se utiliza un conjunto de estándares aceptados por entidades como W3C.

- OpenSearch y SPARQL como lenguajes de consulta
- eGov-SDShare como lenguaje de recolección de datos y descubrimiento de contenidos
- Las especificaciones XML, RDF y OWL para la descripción de recursos y envío de mensajes

## 2.2.2 Arquitectura

iCMS puede operar según dos roles diferenciados, que además son no excluyentes:

- Servidor de contenidos: Un servidor compatible iCMS exporta una interfaz HTTP para el acceso a la información que obtiene de una determinada fuente.
- Consumidor de contenidos: La función de un cliente iCMS es la consulta y recolección de información a través de la interfaz HTTP que un servidor exporta. Un cliente también puede suscribirse a colecciones de un servidor, que le notificará de cambios realizados en ellas.

La implementación de la interfaz desarrollada se distribuye en forma de servidor autocontenido y extensible, puesto que la conexión con sistemas externos se realiza a través de un sistema de drivers. Estos drivers deben cumplir una interfaz que asegura que se soportarán las operaciones expuestas en un punto anterior.

El diseño detrás de iCMS potencia la idea de modularización, proporcionando un sistema basado en plugins que facilita el desarrollo e instalación por parte de terceros de conectores para nuevos tipos de sistemas. Además, permite una mayor claridad y una independencia total entre los distintos sistemas existentes, aislando posibles errores en el sistema. Por otra parte, la implementación de iCMS está escrita en Java, en consonancia con el Marco de Desarrollo de la Junta de Andalucía, pero al definir una interfaz HTTP para las operaciones permite una extensibilidad total, independiente del lenguaje.

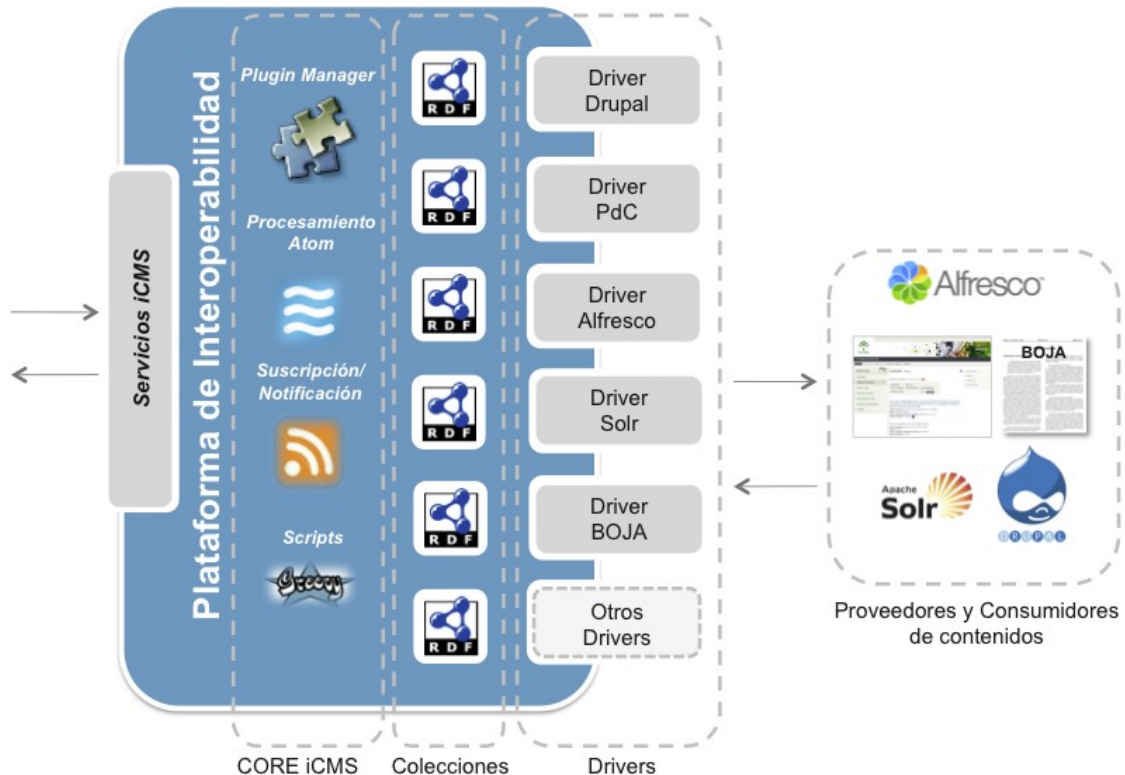


Ilustración 1: Arquitectura de iCMS

## 2.3 Funcionalidad

La Interfaz de Interoperabilidad define, como se ha mencionado, un conjunto de operaciones básico que permite:

- La recolección de los contenidos manejados por un servidor
- La consulta sobre los contenidos de un servidor
- La publicación de nuevos contenidos
- La suscripción a un servidor para recibir notificaciones cuando se producen cambios en los contenidos que gestiona

El uso combinado de estas operaciones permite cumplir con diversos escenarios de mucha utilidad en la actualidad. Pueden identificarse entre otros los siguientes escenarios de uso:

1. Separación entre gestor de contenidos y gestor de presentación: puede utilizarse un producto software distinto para generar y servir a los usuarios las páginas del que se utiliza para gestionar internamente la información, permitiendo así la elección del más adecuado en cada caso. La comunicación entre ambos puede realizarse en tiempo real y con baja latencia mediante iCMS.
2. Sindicación de información contenida en múltiples gestores de contenidos: un único gestor de presentación, encargado de servir a los usuarios las páginas del sitio web, puede construirlas a partir de información almacenada en varios gestores de contenidos distintos y presentar la información de forma integrada.
3. Presentación de información contenida en bases de datos de aplicaciones: frecuentemente es necesario presentar en los sitios web información que reside en sistemas de gestión internos. En lugar de construir sitios web ad-hoc o desarrollar conectores, protocolos y formatos específicos para cada una de estas aplicaciones que accedan a su modelo de datos, con los consiguientes problemas de mantenibilidad, puede emplearse iCMS para proporcionar la información a los servidores de publicación de forma eficiente.
4. Explotación automática de la información contenida en bases de datos de aplicaciones: al exponer la información de un sistema por iCMS, ésta puede utilizarse para cualquier otra aplicación.
5. Web semántica: al utilizar protocolos y formatos estándar de web semántica, puede utilizarse no sólo internamente, sino de cara al exterior, fomentando la reutilización de la información del sector público.

iCMS está diseñado para optimizar estos procesos. Por ejemplo, es posible realizar la sincronización en tiempo real entre servidores, realizando una primera recolección y utilizando el sistema de suscripción y notificación para mantener el sincronismo. También es posible el acceso síncrono de forma muy eficiente a la información gracias a la utilización de las técnicas de cacheo previstas en el protocolo HTTP.

## 2.4 Rendimiento

Durante todo el desarrollo del proyecto se llevaron a cabo una serie de pruebas de stress contra el servidor iCMS para así poder reflejar los tiempos de respuesta y el rendimiento del sistema.

### 2.4.1 Configuración Hardware

Dichas pruebas se realizaron contra un servidor iCMS instalado en una máquina con las siguientes características:

- Procesador Intel Pentium Dual E2200 2,2Ghz
- 2 Gb memoria RAM

La máquina con la que se realizaron las peticiones tenía la misma configuración hardware y estaba instalada dentro del mismo segmento de red que la máquina con el servidor iCMS. Ambas compartían medio con tecnología Ethernet 100BaseTX.

### 2.4.2 Configuración Software

El servidor iCMS corresponde a la versión 0.1.6 con el proveedor del Boletín Oficial de la Junta de Andalucía y corre en un sistema operativo Ubuntu 8.10 (Linux 2.6.27). Por otra parte, la máquina encargada de realizar las peticiones y el análisis del rendimiento tiene la misma instalación de software y hace uso de la herramienta de pruebas de stress Apache JMeter 2.3.4.

El servidor iCMS tiene una cola de conexiones de 100 peticiones y acepta 25 conexiones concurrentes.

### 2.4.3 Prueba 1

Esta primera prueba tiene como objetivo medir el comportamiento del servidor iCMS ante peticiones reiteradas al mismo recurso. Para ello la máquina A realiza 20 peticiones por segundo a iCMS pidiendo los 200 primeros elementos de BOJA. La dirección de petición será <http://maquina:puerto/icms/feeds/boja?items=200>.

La tabla de resultados obtenidos para 1143 muestras es:

#Muestras	Media	Mediana	Min	Máx	%Error	Rendimiento	Kb/sec
1143	3586	3131	273	12863	0,00%	8,2/sec	10386,5

La tabla recoge los datos de rendimiento después de realizar 1143 peticiones a la dirección <http://maquina:puerto/icms/feeds/boja?items=200>. Como se puede observar, el servidor iCMS tiene un rendimiento de 8,2 recursos por segundo y tarda de media 3586 ms en servir los 20 recursos pedidos desde que realiza la petición concurrente hasta que sirve el último. Cabe destacar que no se produce ningún error en las peticiones y que la mediana de tiempos de respuesta está por debajo de la media. El sistema iCMS ha cacheado las respuesta y sólo realizará consultas sobre la caché para despachar las peticiones. Esto ha requerido unas 900 muestras que con una tasa de 20 peticiones cada segundo se llega a 45 segundos de tiempo de estabilización del sistema.

### 2.4.4 Prueba 2

Esta segunda prueba consistirá en simular un entorno de peticiones más realista, pidiendo a iCMS distintas páginas cada vez. Para ello se atacará a las urls <http://maquina:puerto/icms/feeds/boja?items=200>,



[http://maquina:puerto/icms/feeds/boja?items=200&page=2,](http://maquina:puerto/icms/feeds/boja?items=200&page=2)  
[http://maquina:puerto/icms/feeds/boja?items=200&page=3,](http://maquina:puerto/icms/feeds/boja?items=200&page=3)  
[http://maquina:puerto/icms/feeds/boja?items=200&page=4,](http://maquina:puerto/icms/feeds/boja?items=200&page=4)  
[http://maquina:puerto/icms/feeds/boja?items=200&page=5.](http://maquina:puerto/icms/feeds/boja?items=200&page=5)

Al igual que en la primera prueba, se realizarán 20 peticiones por segundos.

La tabla de resultados obtenidos para 1868 muestras es:

#Muestras	Media	Mediana	Min	Máx	%Error	Rendimiento	Kb/sec
1868	37	38	2	154	0,30%	517,5/sec	10386,8

Se observa una amplia mejora en el rendimiento comparándola con la tabla de resultados de la primera prueba. Esto es debido en parte a que en esta segunda prueba, la mayoría de los recursos están ya cacheados de la primera prueba.

### 2.4.5 Valoración final

Los resultados obtenidos son muy satisfactorios, haciendo especial hincapié en el bajo porcentaje de errores del sistema ante una carga moderada y en un tiempo de respuesta más que aceptable.

En todo caso, como en cualquier infraestructura de red, estos resultados siempre vendrán limitados por la velocidad de la red en donde se encuentre el servidor y el cliente de peticiones.

## 2.5 Próximos hitos

Actualmente el proyecto se encuentra ya disponible en el Catálogo de Software de la Junta de Andalucía, para su libre implantación en los distintos organismos de la Junta de Andalucía, aunque se han planteado una serie de mejoras para incrementar su funcionalidad. Entre las mismas podemos destacar:

- ampliaciones al protocolo de comunicación y mejoras de mantenibilidad de los sistemas.
- integración con un backend de tipo SPARQL para utilizar toda la potencia que proporciona este lenguaje de consulta.
- nuevos drivers para permitir la interoperabilidad con otras fuentes de datos presentes en la Junta de Andalucía, tales como *OpenCMS* y cualquier fuente de datos SQL genérica.