

LA CIENCIATK DEL CSIC. UN PORTAL MULTIMEDIA CON TECNOLOGÍA J2EE PARA LA DIFUSIÓN DE LA CULTURA CIENTÍFICA EN INTERNET

Autores.

Clara Cala Rivero

Centro Técnico de Informática
Consejo Superior de Investigaciones Científicas

Ángel L. Rodríguez Alcalde

Centro Técnico de Informática del CSIC
Consejo Superior de Investigaciones Científicas

Gil Martín Estrada

Centro Técnico de Informática del CSIC
Consejo Superior de Investigaciones Científicas

RESUMEN

Cienciatk es un proyecto del CSIC para el fomento y difusión de la cultura científica, que entrará en producción a finales de 2.007. Facilita a través de Internet y mediante un portal Web el acceso a casi 40.000 documentos multimedia de carácter multidisciplinar (vídeos, imágenes y sonidos) de los fondos de la Mediateca del CSIC.

La comunicación aborda distintos aspectos del proyecto: tecnología empleada, funcionalidad ofrecida por el portal, acciones para creación de una imagen propia y los mecanismos introducidos para la protección de los derechos de autor y productoras.

Se describe el Framework utilizado para el desarrollo del portal, basado en J2EE y Spring-JSF-Hibernate, así como la arquitectura hardware, sistema de streaming y condiciones de las emisiones

LA CIENCIATK DEL CSIC. UN PORTAL MULTIMEDIA CON TECNOLOGÍA J2EE PARA LA DIFUSIÓN DE LA CULTURA CIENTÍFICA EN INTERNET

Cienciatk es un proyecto del Consejo Superior de Investigaciones Científicas -CSIC- orientado al fomento y divulgación de la cultura científica mediante la difusión de distintos fondos multimedia de carácter multidisciplinar que pertenecen a colecciones propias del CSIC o que, si no pertenecen a este organismo, están disponibles en su Mediateca.

El acceso a los fondos se realizará a través de Internet y, en concreto, desde un portal Web de acceso público que permite hacer búsquedas por distintos criterios y cuenta con varios canales temáticos con programación diaria y continua.

La primera fase del proyecto, cuya entrada en producción está prevista para el último trimestre de 2.007, permitirá el acceso público a aproximadamente 3.500 vídeos, 20.000 imágenes y 200 sonidos.

Este documento refleja la génesis del proyecto, así como distintas fases y aspectos del mismo, entre los que se encuentra:

- Definición, diseño, desarrollo del portal.
- Entorno tecnológico: Framework y arquitectura hardware.
- Proceso para la codificación de vídeos y sistema de streaming empleado.
- Creación de una imagen propia –marca con logotipo-, que se ha protegido mediante la reserva comunitaria de la marca y reserva de los dominios de Internet.
- Estudio y resolución de la problemática de los temas derivados de la protección de los derechos de autor.

Origen del proyecto. La MEDIATECA del CSIC

El proyecto Cienciatk tiene su antecedente en la Mediateca del CSIC (<http://www.csic.es/mediateca/mediateca.html>).

Mediateca surge de la firma de un convenio entre el Consejo Superior de Investigaciones Científicas (CSIC), la Comunidad de Madrid y la Asociación Española de Cine Científico (ASECIC), por el que las instituciones se comprometen a conservar y difundir, y en una segunda etapa promover la creación de nuevos productos multimedia de contenido científico.

Tras su inicio, en mayo de 2003 ha



permitido a un elevado número de personas y grupos, normalmente escolares e investigadores, la consulta de sus colecciones de vídeos, imágenes y sonidos. El acceso a los fondos se realiza exclusivamente desde los 28 terminales ubicados en las salas de la Mediateca, ubicadas en el Museo de Ciencias Naturales de Madrid.

Los fondos iniciales de la Mediateca procedían de la Asociación Española de Cine Científico y de producciones de los diferentes Institutos del CSIC. A éstos se han unido en estos años los de WWF_ADENA, CIEMAT, UNED, Videoteca Científica UNICAJA, VIDEOMED, TELENATURA, VIDEOPRIMARIA, CENEAM, Videoteca Científica CAI, etc.

Ante el éxito del proyecto se planteó la conveniencia de facilitar el acceso a los fondos audiovisuales de Mediateca a través de un canal público, como es Internet. Esto dio lugar al proyecto **Cienciatk**, si bien este proyecto tiene como segundo objetivo el potenciar la producción y difusión de fondos desarrollados por centros y equipos de investigación en los que participa el CSIC.

ELEMENTOS DEL PROYECTO

Los destinatarios

Cómo se ha señalado anteriormente, el proyecto tiene como objetivo ofrecer acceso público y gratuito –en abierto- a todos los fondos que no estén sujetos a alguna limitación o restricción impuesta por autores o productoras que impida su emisión o visionado a través de Internet, es decir, tiene vocación universal.

Adicionalmente, se estudia la posibilidad, para una segunda fase, de firmar acuerdos con instituciones educativas o dedicadas a la investigación, prioritariamente Universidades, a las que se facilitará el acceso en condiciones de alta calidad tanto a los fondos de carácter público como a aquellos “restringidos” para los que se disponga de autorización para este tipo de emisión.

La posibilidad de ofrecer un servicio de alta calidad se ve favorecido, en este caso, por el hecho de que las instituciones académicas habitualmente tienen como proveedor de servicios de comunicaciones a RedIris, que ofrece ancho de banda suficiente y servicio multicast.

Los Fondos

La colección actual de la Mediateca del CSIC que se desea poner en Internet, a través del proyecto Cienciatk, supera los 3.500 vídeos, 20.000 imágenes y 200 sonidos, que abarcan todas las disciplinas científicas, técnicas y humanísticas.

El fondo tiene distintas procedencias, pudiendo catalogarse por su origen como:

- Colecciones propias del CSIC procedentes de sus archivos, museos, ...
- Resultados de proyectos realizados por grupos de investigación del CSIC o por grupos mixtos del CSIC y otras instituciones o entidades

- Vídeos o películas de productoras, procedentes de acuerdos con la Asociación Española de Cine Científico o de acuerdos realizados directamente por Mediateca.
- Vídeos o películas de certámenes de cine científico.

Actualmente, se dispone de autorización para el visionado del material en la sala de la Mediateca que el CSIC posee en el Museo de Ciencias Naturales de Madrid. Puesto que el proyecto Cienciatk modifica sustancialmente el alcance de los acuerdos firmados, es necesario realizar la revisión de éstos con cada autor o productora a fin de conseguir su autorización para la emisión del material a través de Internet.

Este trabajo se está llevando a cabo y, en el momento de la redacción de esta comunicación, se desconoce el número exacto de documentales que podrán visualizarse en Internet, hecho que influye en la fecha final de arranque del proyecto.

Personal dedicado al proyecto

El equipo dedicado al proyecto depende del Área de Cultura Científica y del Centro Técnico de Informática del CSIC y está compuesto por:

- El coordinador general de la Mediateca
- Dos documentalistas
- Un especialista en imagen y sonido y un técnico de apoyo
- Equipo técnico de informática y comunicaciones.
- Empresa contratada para el desarrollo del portal.

Considerando el elevado número de vídeos que se reciben periódicamente y el tiempo estimado en recepcionar, visionar, documentar, digitalizar e introducir en el sistema cada uno de ellos, se está revisando la composición del equipo, planteándose la opción de incrementarlo.

La tarea que presenta mayor déficit de personal es la relativa a la catalogación, planteándose en estos momentos la posibilidad de externalizar esta tarea o incrementar el número de documentalistas adscritos al proyecto. La tarea de documentación es sumamente compleja debido a la diversidad de áreas temáticas del material que se recibe y a que existen dos tipos de destinatarios bien diferenciados: el ciudadano, en general, y el investigador o especialista en una materia.

Creación de una imagen propia

Desde el inicio del proyecto, se determinó la importancia de dotar a Cienciatk de una denominación y logotipo propio.

Nombre. Tras valorar distintas posibilidades, se seleccionó como nombre del proyecto “**CIENCIATK**”, ya que:

- Hace referencia a la Ciencia y a la Colección
- Hace referencia al medio más importante o representativo de los medios audiovisuales que se ofrecen: el vídeo
- Compone un acrónimo muy usado por los jóvenes en sus comunicaciones habituales por SMS, Chats e Internet

Logotipo. Características

- Se ha diseñado a una sola tinta: sencillo de duplicar
- Es visible y reconocible casi a cualquier tamaño –desde los más pequeños a los más grandes-
- Admite cualquier color
- Coherente con el logo, existe espacio para ubicar los correspondientes a patrocinadores o entidades relacionadas



Medidas de protección de la identidad del proyecto

Se han adoptado las siguientes medidas para la protección de la imagen seleccionada:

Reserva de los **nombres de dominio** en Internet: www.cienciatk.es, www.cienciatk.org y www.cienciatk.com

La reserva de estos dominios tiene por finalidad evitar un uso fraudulento, malintencionado o perjudicial para la imagen del proyecto, aunque el dominio propio es: www.cienciatk.csic.es

Reserva de la **marca**. Se ha procedido a solicitar a la OAMI, Oficina de Armonización del Mercado Interior, la marca CIENCIATK (c/logotipo) como marca comunitaria para las clases 9, 41 y 42 para garantizar los siguientes productos o servicios:

Clase 9. Software y programas informáticos; publicaciones en formato electrónico; software informático y publicaciones en formato electrónico suministrados en línea desde base de datos o desde puntos de Internet (inclusive sitios web).

Clase 41. Educación; formación; esparcimiento; actividades deportivas y culturales; publicación de información visualizable electrónicamente en forma de textos, gráficos, imágenes y sonido, y accesible a través de redes de datos; edición y edición electrónica.

Clase 42. Servicios científicos y tecnológicos así como servicios de investigación y diseño relativos a ellos; servicios de análisis y de investigación industrial; diseño y desarrollo de ordenadores y software; servicios de información en línea en el ámbito de la investigación científica y tecnológica procedente de bases de datos informáticas o de Internet; creación y explotación de una base de datos; creación de programas para el tratamiento de la información; creación de sitios web en Internet; servicios de

investigación e información informatizados en el ámbito de la investigación científica y tecnológica.

ENTORNO FUNCIONAL. SERVICIOS

El sistema de información de Cienciatk integra el **portal Cienciatk** en Internet y un sistema de back-office o **aplicación de administración**.

El portal Cienciatk. Principales Funcionalidades

El portal Cienciatk es una Web pública que permite el acceso a los fondos multimedia a través de Internet. Dispone de las siguientes funcionalidades:

The screenshot displays the Cienciatk website interface. At the top, there is a header with the logo 'CIENCIATK Consejo Superior de Investigaciones Científicas' on the left, a central banner for 'AÑO CIENCIA 2007' with '365 días de Desarrollo' and the CSIC logo on the right, and a search bar labeled 'BUSCADOR DE CONTENIDOS' with a 'Buscar' button and a 'BÚSQUEDA AVANZADA>>' link. Below the header, there are several sections: 'FONDO DOCUMENTAL' with links to 'Acerca del fondo documental', 'Listado de categorías', and 'Novedades'; 'CIENCIATK' with links to '¿Conoces cienciatk?', 'Atención usuario', and 'Info. aportaciones'; and 'PROGRAMACIÓN DE EMISIONES' which includes a calendar for October 2007. The main content area is titled 'EMISIONES ON-LINE' and features four thematic channels: 'CIENCIA Y TECNOLOGIA' (PROGRAMA AUTOPIA, DEMOSTRACIÓN PARA NIÑOS), 'HUMANIDADES Y CIENCIAS SOCIALES' (Crónica Universal de Nuestra Época: 3. LA GRAN GUERRA (1914-1918)), 'EVENTOS Y ACTOS PUBLICOS' (AUDIOVISUAL DEL MNCH. EXPOSICIÓN EL RETORNO), and 'CIENCIAS DE LA VIDA' (Residuos domésticos: SISTEMAS DE GESTIÓN DE LOS RESIDUOS DOMÉSTICOS). Each channel includes a video player icon, the language 'Español', the start time, and links for 'Ver ficha »»' and 'Ver Video »»'. At the bottom, there is a footer with technical specifications: 'Resolución mínima: 1024x768 Navegadores: Firefox 1.5/Internet Explorer 6.0' and 'Todas las imágenes, vídeos y audios están protegidas por sus respectivos derechos de autor', along with logos for the Spanish Government and CSIC.

Canales temáticos

La zona central de la pantalla principal del portal presenta cuatro canales temáticos públicos destinados respectivamente a Humanidades y Ciencias Sociales, Ciencias de la Vida, Ciencia y Tecnología y Eventos y Actos Públicos.

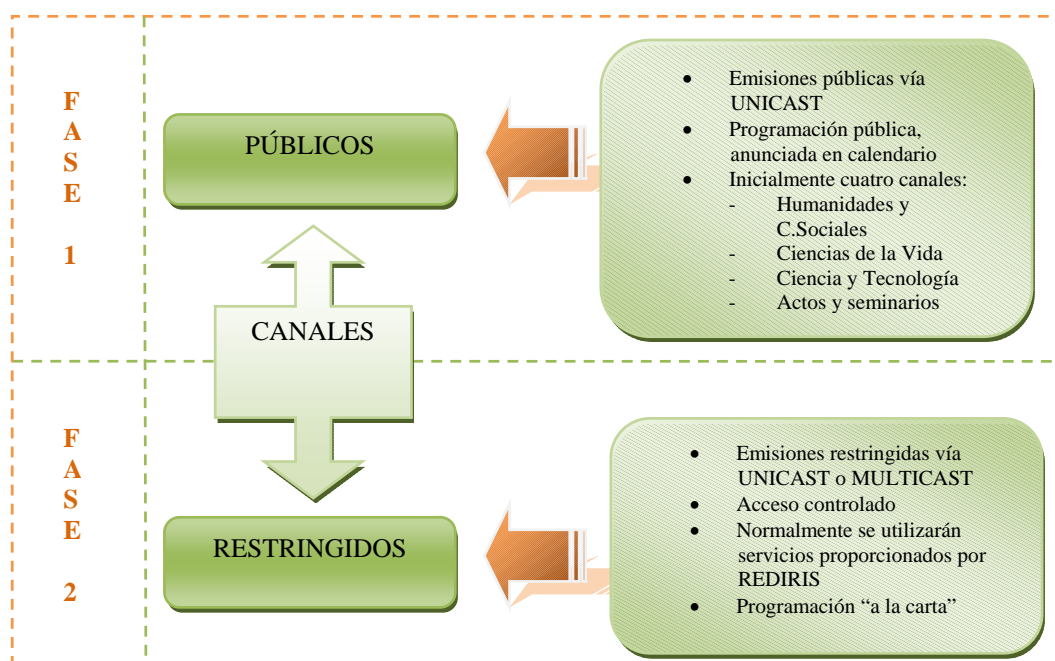
Estos canales emiten de manera continuada. La programación de la parrilla se realiza a través de la aplicación de administración o back-office ya mencionado. Aunque el número de canales por defecto se ha fijado en cuatro, el sistema permite configurar y, por consiguiente, presentar cualquier otro número de canales de manera automática.

Los canales presentan aquellos vídeos de mayor actualidad, calidad, interés por su temática o que hayan sido previamente solicitados por instituciones para algún evento.

La programación se puede consultar en el calendario de programación de emisiones que ofrece la propia Web.

El método de emisión utilizado es UNICAST, que permite la emisión de un emisor a un receptor, esto quiere decir que se estarán realizando tantas emisiones de un mismo vídeo como receptores se encuentren viéndolo simultáneamente.

En una segunda fase, se prevé la creación de canales “restringidos”, que ofrecerán emisiones “a la carta” destinadas a instituciones de carácter científico o educativo con la que se disponga de acuerdos ad-hoc.



El método de emisión a emplear en la distribución de vídeo a través de los canales restringidos será UNICAST o MULTICAST.

Se utilizará MULTICAST cuando se desee emitir un vídeo a distintos destinatarios simultáneamente, siendo este tipo de emisión la estrategia más eficiente, ya que permite realizar un único envío, con lo que el ancho de banda empleado es óptimo.

La posibilidad de utilizar la estrategia MULTICAST para ofrecer streaming es factible con las instituciones académicas y científicas, al ser éste, un servicio ofrecido por RedIris (Red.es), proveedor de comunicaciones común a estas instituciones, mientras que otros operadores de comunicaciones no siempre lo ofertan.

Sistema de Búsquedas y Resultados

Cienciatk dispone de un buscador de contenidos y un sistema de búsqueda avanzada, teniendo como motor de búsqueda Lucent.

El buscador avanzado permite realizar búsquedas combinadas por distintos campos, como título, año de producción, formato, título original, nombre propio, nombre vulgar, comunidad autónoma, provincia, país, año de producción y formato

El resultado de la consulta se presenta en una pantalla que muestra todos los resultados que cumplen las condiciones introducidas. Los datos que aporta esta pantalla son:

- Un icono representativo del tipo de material: vídeo, imagen o sonido
- Una imagen reducida, en caso de fotografías o un fotograma del vídeo, su título y signatura. En el caso de los vídeos se indica asimismo el idioma, idioma de subtítulos si están disponibles y duración.
- Las opciones de acceso al material y a la ficha técnica.

The screenshot shows the CIENCIATK website interface. The main content area displays search results for 'BÚSQUEDA AVANZADA'. The results are organized into a grid of six items. Each item includes a thumbnail image, a title, a signature number, language, subtitle availability, and duration. For example, the first result is 'MITOSIS: CARIOKINESIS Y CITOKINESIS EN CÉLULAS ANIMALES Y VEGETALES.' with a signature of 2281, German language, and a duration of 16:0 min. The page also features a sidebar with navigation options and a calendar for October 2007.

En el caso de vídeos, el usuario tiene la opción de ver un previo de 30 segundos, ver el vídeo completo o acceder a la ficha técnica. En el resto de los casos pueden visualizar a tamaño original la imagen o escuchar el sonido, así como acceder a la ficha técnica.

Por defecto, los resultados obtenidos se muestran en grupos de seis, aunque el usuario puede configurar este valor.

Visualización de Previos y Vídeos. Ficha

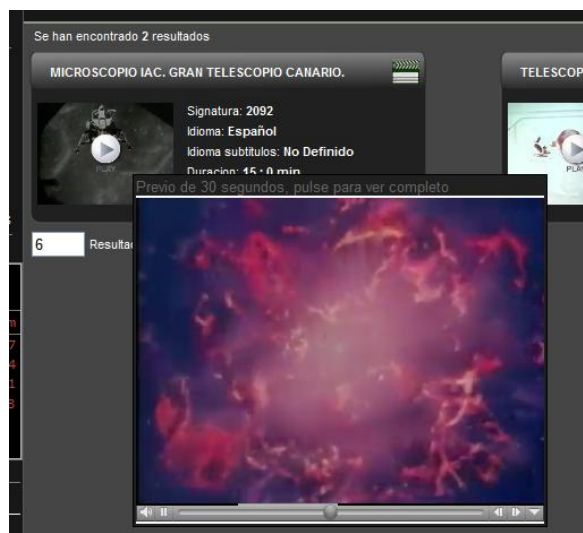
Se ha establecido un sistema, mediante la opción **Ver Previo** que permite al usuario aproximarse al contenido y calidad de cualquier vídeo mediante un corte de 30 segundos.

Los previos se generan de manera automática. Para ello, se ha desarrollado un script que durante el proceso de digitalización y conversión de formato extrae un fragmento del vídeo de 30 segundos, haciendo un corte entre los minutos 5':00" y 5':30". La imagen que identifica cada vídeo corresponde al primer fotograma del corte mencionado.

En el sistema de archivos, queda pues almacenado el vídeo, el previo del vídeo y el fotograma que lo identifica.

Un previo se activa cuando se pasa con el puntero del ratón por encima del fotograma que muestra la pantalla. Para ver el vídeo hay que hacer click sobre dicha imagen.

El sistema ofrece asimismo la posibilidad de ver la **ficha técnica** de cada uno de los vídeos, imágenes o sonidos. Dicha ficha contiene información relativa a título, autor, idioma, idioma de los subtítulos, duración, taxonomía (reino, filo, clase, orden, familia, género, especie), localización geográfica (comunidad autónoma, provincia, país), fondo, país de producción, observaciones, etc.



ELEMENTOS TÉCNICOS DEL PROYECTO

El desarrollo e implantación de la primera fase del proyecto Cienciatk, se ha llevado a cabo en un periodo aproximado de un año y con un coste total aproximado de €160.000, durante el cual se han abordado las siguientes actividades y tareas:

- Definición, análisis, diseño, desarrollo y puesta en marcha del portal y de la aplicación de gestión.
- Adquisición e instalación de la plataforma de sistemas: entornos de desarrollo, preproducción y producción.
- Recodificación de vídeos a formato H.264 y su disposición para ser emitidos mediante el protocolo RTSP
- Migración de base de datos

El desarrollo del portal y aplicación de gestión se ha realizado con la colaboración de una empresa externa. Ha participado en la fase de requisitos funcionales el personal de Mediateca y del Centro Técnico de Informática, unidad que se ha encargado asimismo del soporte y equipamiento de sistemas.

En esta fase del desarrollo se ha llevado a cabo la migración de la base de datos de Mediateca sobre PostgreSQL a la nueva estructura de la base de datos creada para el proyecto y que tiene como motor de base de datos Oracle 10g.

Se han desarrollado asimismo los scripts que han permitido generar los cortes o fragmentos de los vídeos de 30 segundos y los fotogramas que sirven de portada, la definición y configuración de las estaciones de digitalización para conversión de más de 3500 vídeos, de distintos formato a formato H.264/MPEG-4 AVC.

La segunda fase del proyecto, que se abordará próximamente, intentará optimizar el código empleado para alcanzar el máximo nivel de accesibilidad posible, considerando la dificultad que subsiste en un proyecto de estas características, cuyo objetivo es poner a disposición de los usuarios documentos multimedia.

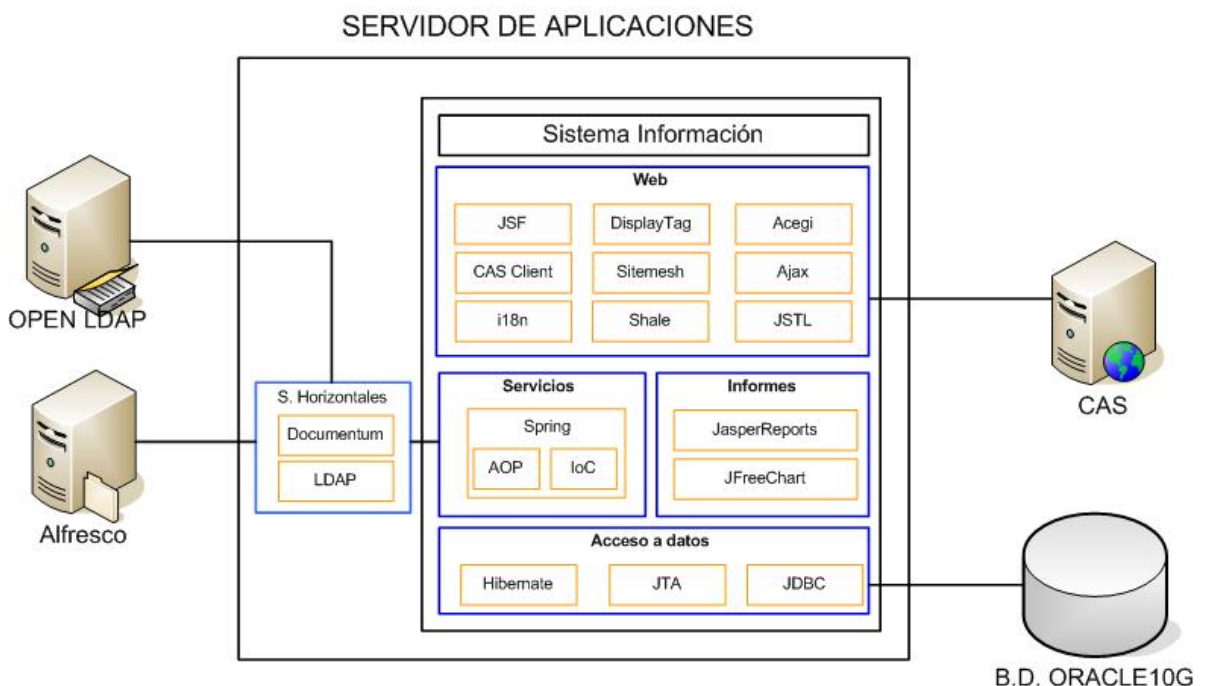
Asimismo, en dicha fase, se potenciará la cooperación con otras instituciones, fundamentalmente del mundo académico y científico, así como la visibilidad del trabajo realizado por los distintos grupos de investigación y centros del CSIC. Incluirá asimismo un mayor número de funcionalidades, como la posibilidad de solicitud de peticiones “a la demanda” y creará nuevos de perfiles de acceso distintos de los tres actualmente existentes: usuario Internet, usuario Intranet del CSIC y administrador.

ENTORNO TECNOLÓGICO

En el ámbito de la tecnología empleada, el desarrollo se ha realizado utilizando el Framework del CSIC (se puede encontrar información en el documento “El Framework de desarrollo del Consejo Superior de Investigaciones Científicas, presentado a Tecnimap_2006

http://www.csae.map.es/csi/tecnicmap/tecnicmap_2006/01T_PDF/el%20framework.pdf).

De acuerdo al Framework, cuyo esquema general muestra la figura siguiente, el sistema utiliza una arquitectura **J2EE** en la que se integran básicamente **Spring-JSF-Hibernate**, usando **Maven** como la herramienta de gestión del proyecto a partir de un fichero de configuración XML. El motor de búsquedas es **Lucene**.

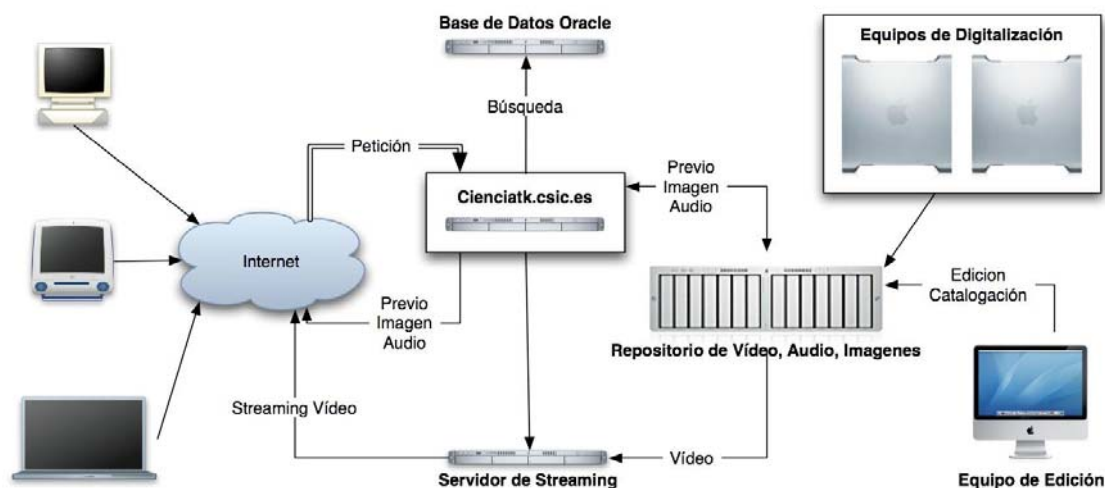


Existe conexión con el LDAP del CSIC –OPEN LDAP- para la autenticación y el single sign-on se realiza a través de CAS.

Arquitectura tecnológica

La plataforma que da soporte al proyecto y que muestra la siguiente imagen está basada en:

- Servidor de Aplicaciones: JBoss
- Base de Datos: Oracle 10g
- Servidor de streaming Apple xServe Doble procesador G5, 2 GB RAM, 1,5 Tb Hard disk y 2 GigaEthernet. Sistema operativo MacOS X 10.4 y software de streaming es Quicktime Streaming Server
- Almacenamiento. NAS de EMC -3TeraBytes-
- Equipos de digitalización y edición, basados en el software EPISODE PRO.



Condiciones técnicas de las emisiones

Las emisiones se realizan mediante el método UNICAST, aunque para la segunda fase se prevé dar un servicio MULTICAST a Universidades y otros centros de investigación o educativos.

Se ha incluido en vídeos e imágenes una **marca de agua** que contiene el logotipo de Cienciatk, marca que se introduce durante el proceso de digitalización.

Respecto de la **Protección** frente a copias, el protocolo de transmisión RTSP - *Real Time Streaming Protocol*- empleado para la difusión de los vídeos dificulta al usuario final conservar una copia local del material al que accede.

Formato de vídeo. En esta primera fase, el sistema ofrece los vídeos en calidad alta, a partir del formato de codificación H.264/MPEG-4 AVC, lo que permite una codificación de alta compresión con una buena calidad de imagen.

Para la próxima fase se espera introducir tres niveles de calidad o resolución para los vídeos, facilitando al usuario final seleccionar la resolución más adecuada al ancho de banda del que disponga.

Ancho de banda estimado. Se considera que cada usuario consume un ancho de banda medio de 700 Kb. Al ser un sistema de nueva implantación no se dispone de estimaciones realistas sobre la posible demanda del mismo.

Por este motivo, se ha diseñado una arquitectura escalable a nivel de servidores, que permitirá ofrecer un crecimiento adecuado a las demandas.

Por otra parte, el CSIC dispone de un elevado ancho de banda de conexión con Internet, por lo que inicialmente no se prevén problemas en este apartado; a pesar de ello, se han analizado mecanismos alternativos en caso de que éstos se produzcan, como priorización de tráfico y/o externalización de la distribución del contenido.

CONCLUSIONES

Cienciatk es un proyecto de nueva creación, definido para su implantación por fases. La primera fase estará en producción en el último trimestre de 2.007.

Las principales dificultades o mayores incertidumbres en la puesta marcha del proyecto no han sido de tipo técnico, sino las derivadas de la necesidad de establecer mecanismos que garanticen la protección de los derechos de los autores y productoras –consultas legales, definición de procedimientos y recogida de autorizaciones e implantación de mecanismos anti copia-.

Tras la puesta en marcha de la primera fase, se espera abordar una nueva versión del sistema con mayor número de funcionalidades y que facilite su integración con herramientas estándares de documentación y gestión de fondos multimedia y con el sistema Open Access del CSIC, al tiempo que debe aumentar la cooperación con otras instituciones.

La tecnología empleada será utilizada para la emisión de actos y eventos del CSIC por Internet mediante vídeo-streaming.