

BIBLIOTECNICA 2.0 : LA NUEVA BIBLIOTECA DIGITAL DE LA UPC

Jordi Prats, Toni Prieto, Oriol Rico, Jordi Serrano-Muñoz

Servei de Biblioteques i Documentació. Universitat Politècnica de Catalunya
jordi.prats@upc.edu antonio.juan.prieto@upc.edu oriol.rico@upc.edu j.serrano@upc.edu
http://bibliotecnica.upc.edu

Resumen/ Abstract: Se expone la tarea realizada por el Servei de Biblioteques i Documentació de la Universitat Politècnica de Catalunya para mejorar el portal web del servicio aplicando conceptos de la web 2.0. Un modelo de gestión completamente descentralizado, y la integración de productos hasta la fecha tratados de forma separada como el repositorio institucional (UPCommons) i la base de datos bibliográfica es considerado uno de los aspectos mas relevantes de el nuevo portal del las bibliotecas de la UPC

Palabras clave / Key words: DRUPAL, gestión de contenidos, Bibliotecas universitarias, MILENIUM, UPCommons, repositorios, Universitat Politècnica de Catalunya, UPC, web 2.0, content management, repositories, university libraries

INTRODUCCIÓN

El primer portal de las bibliotecas de la Universitat Politècnica de Catalunya se remonta al curso 1994-1995 (1), en donde cada biblioteca de centro organizaba sus contenidos de forma completamente descentralizada,

aunque con el denominador común del acceso al OPAC y a algunos recursos digitales. El resultado fue un crecimiento desigual, generalmente condicionado por las necesidades del centro y también según las capacidades de las personas responsables de su mantenimiento.

A partir del año 2000, un proceso de reflexión y organizativo, reforzado por el crecimiento de los contenidos digitales, nos lleva al actual Bibliotecnia (2) en donde se homogenizan los contenidos y la imagen corporativa de las bibliotecas UPC.

Durante estos años, Bibliotecnia, ha ido creciendo y a la vez se han añadido nuevas piezas como UPCommons, el portal de conocimiento de la UPC (<http://upcommons.upc.edu>), la Factoría de recursos docentes (<http://www.upc.edu/factoria>) o FenixDoc el portal de la producción científica (<http://bibliotecnia.upc.es/FenixDoc/>).

En los últimos meses dentro del plan estratégico de la Bibliotecas UPC APREN 2010 (3) se abre un proceso de reflexión respecto a los aspectos anteriores, así como de qué modo afrontábamos el nuevo espacio europeo de educación superior (EEES). Se llegó a la conclusión de que era necesario una reorganización de los contenidos y servicios, así como que la gestión de estos se pudiera realizar de forma completamente descentralizada y adaptada a las necesidades y características de cada centro, teniendo en cuenta aspectos como la posibilidad de crear flujos de trabajo en el proceso de creación y edición de contenidos, la posibilidad de implementar y utilizar herramientas 2.0 y que los usuarios puedan personalizar sus contenidos.

HACIA UN NUEVO PORTAL

Se crearon tres grupos de trabajo relacionados con la estructura y contenidos, con la usabilidad y accesibilidad, y con el desarrollo tecnológico que debería tener el nuevo portal al que denominamos “Bibliotecnia 2.0”.

De los dos primeros grupos, solo comentar la encuesta realizada entre usuarios activos y no activos de diferentes bibliotecas, y el trabajo de benchmarking que no solo incluyó bibliotecas, sino también portales de prensa y populares portales como Amazon o Ebay entre otros.

El grupo de desarrollo tecnológico, se planteó la definición de:

- El modelo de desarrollo TIC
- Las herramientas tecnológicas prioritarias
- El modelo para el desarrollo de proyectos TIC
- Estándares y los modelos de integración i interoperabilidad prioritarios

Entre otras actividades, se evaluaron diferentes CMS o gestores de contenidos tanto propietarios como de código abierto:

- Vignette (<http://www.vignette.com>)
- Documentum (<http://www.documentum.com>)
- Plone (<http://www.plone.org>)
- Joomla! (<http://www.joomla.org>)
- Drupal (<http://drupal.org>)
- Alfresco: (<http://www.alfresco.org>)
- uPortal (<http://www.uportal.org>)
- Typo3 (<http://typo3.com/>)
- eBD: (<http://www.oasyssoft.com>)

Los requisitos finales para la selección del CMS, fueron:

- Código abierto
- PHP
- Base de datos relacional
- Comunidad de usuarios importante
- Documentación disponible
- Autenticación LDAP
- SSL
- XHTML
- Soporte a la gestión multiidioma
- Editor WYSIWYG
- Corrector ortográfico
- Plantillas para el diseño de las páginas
- Gestión de imágenes o la posibilidad de poder vincularse a un banco de imágenes
- Soporte recomendaciones WAI

De los CMS evaluados, finalmente nos centramos en tres: Joomla!, Drupal y Typo3. Analizadas las prestaciones, en aquel momento (agosto

2008 Typo3 nos parecía que se adaptaba mejor a nuestros requisitos (Figura 1)



Figura 1. Que ofrecen los tres CMS seleccionados

La constante evolución de estos productos con nuevas versiones, módulos, etc.. (4) y las alertas establecidas para mantenernos informados de esta evolución como el estudio de casos de Lee Eden Bradford (5), nos hizo cambiar de opinión en la selección del CMS en el corto período de tiempo en que se debatieron los contenidos que debía tener el nuevo portal Bibliotecna 2.0. Una nueva versión de DRUPAL y la evolución de sus módulos, especialmente la autenticación y la gestión de roles y grupos nos obligó a revisar nuestra decisión, (Figura 2).

Aunque los dos productos cumplían los mismos requisitos, la decisión final estuvo condicionada por la facilidad de instalación y configuración y por tanto un “Start-Up” en donde la evolución y los resultados del site eran plausibles de modo casi inmediato, un segundo factor fue la documentación y la vitalidad de la comunidad de usuarios de DRUPAL, el tercer motivo fue la rápida curva de aprendizaje que debíamos exportar a los diferentes colectivos o roles de las bibliotecas de la UPC.



Figura 2. Resultado en la segunda revisión

¿POR QUÉ UN GESTOR DE CONTENIDOS?

El término WEB 2.0 acuñado por Tim O'Reilly (6), que hace partícipe de los contenidos a los usuarios a la vez que los dueños de la información que depositan, ha sido el elemento inspirador de este nuevo modelo de biblioteca y servicios digitales.

Por ende, era necesario dotar al portal de las bibliotecas UPC de herramientas y utilidades que facilitaran la extracción e introducción de información de modo sencillo y claro, aunque respetando para ello todos los estándares existentes, abrirlo al usuario y de que este sea capaz de poder realizar todas las acciones únicamente con la ayuda de un navegador.

La implementación de un nuevo OPAC como Millennium, así como su biblioteca digital, requería un profundo cambio en todos sus sistemas de información que realiza una reorientación del producto centrándolo en servicios hacia el usuario.

Pero llevar a término una tarea como la reorientación de la biblioteca digital, en la que se entrelazan diferentes elementos como el catálogo, recursos electrónicos, depósitos institucionales, servicios, etc. hizo necesario disponer de una herramienta que sirviera para englobar, interrelacionar y

ofrecer servicios transversales a todos ellos, servicios, por otro lado, que facilitaran convertir nuestro web en un entorno 2.0.

La conclusión ha sido que esta tarea se debe dejar en manos de un gestor de contenidos.

¿POR QUÉ DRUPAL?

Se ha expuesto anteriormente, pero las principales características que hacen a DRUPAL ideal para ese cometido ha sido su facilidad de lenguaje de implementación, con conocimientos técnicos se puede implementar y mantener un sitio web. La existencia de una potente comunidad de desarrollo también nos facilita disponer del sitio constantemente actualizado, y proveer diferentes funcionalidades que se obtienen de los módulos que se publican en el sitio oficial de DRUPAL.

En paralelo, la posibilidad de disponer de múltiples taxonomías para categorizar contenidos, ofertar el “site” en varios idiomas, así como sistemas para añadir comentarios son algunas de las características nativas a DRUPAL, han sido aspectos que a priori nos han facilitado la implementación de servicios 2.0.

UNA DISYUNTIVA: ¿QUÉ HACER CON LAS OTRAS APLICACIONES?

La tarea más tediosa, a priori, puede resultar la importación y explotación de los recursos que se encuentran fuera de la plataforma, como pueden ser el catálogo o los depósitos institucionales, cada uno de ellos con sus propios protocolos de comunicación.

Es en este punto en el que el implementador debe plantearse si debe catalogar aquellos recursos que se encuentran en un depósito institucional e interconexionar únicamente el catálogo (MILLENIUM) con DRUPAL o bien dejar a manos del gestor de contenidos la implementación de los protocolos de importación de los metadatos de ambas fuentes.

Ambas opciones tienen sus ventajas e inconvenientes, que intentaremos repasar a continuación.

La primera de ellas, catalogar absolutamente todo, impide tener cierto control sobre que partes del catálogo se quiere mostrar en el sitio web, dado que el protocolo Xrecord¹ no permite filtraje de contenidos en base a campos del registro, por no hablar de los costes de catalogación directamente proporcionales a la cantidad de registros encontrados en los depósitos institucionales. Remarcar también que la disponibilidad de un cierto elemento se verá influenciada por el tiempo necesario para su catalogación y la imposibilidad de mantener cierto control sobre elementos constantemente en cambio, como por ejemplo, una página web.

No obstante esta opción puede resultar interesante si se quiere disponer de una única fuente de datos, aunque quizá en ese caso DRUPAL pueda suponer una herramienta demasiado grande para tal cometido.

La segunda opción, era implementar los motores de interconexión en el propio gestor de contenidos, esta tarea, puede resultar más laboriosa inicialmente, pero a la larga aporta una serie de beneficios. Mayor control sobre los contenidos importados e inmediatez en el sitio web una vez publicado un determinado recurso en el depósito son solo algunas de las ventajas. No hay que perder de vista que las taxonomías deben ser las mismas que en el catálogo, requerimiento no demasiado difícil de cumplir si se tiene en cuenta que en la mayoría de sitios los metadatos son validados contra el mismo catálogo y revisados por el propio bibliotecario.

Finalmente en el Servei de Biblioteques i Documentació de la UPC, optamos por hacer una estrategia mixta. La segunda opción ofrecía mayor control sobre los contenidos importados, mientras que una catalogación de aquellos contenidos de los depósitos que ya disponen de su versión en papel se presentaba como una opción razonable y a bajo coste, entre otras cosas porque es algo que ya se realiza antes de la planificación de este nuevo proyecto. En la figura 3 se puede observar cual será la estrategia general a seguir.

¹Xrecord es el servicio web que ofrece el catálogo Millennium para poder consultar y extraer datos de un cierto registro del catálogo.

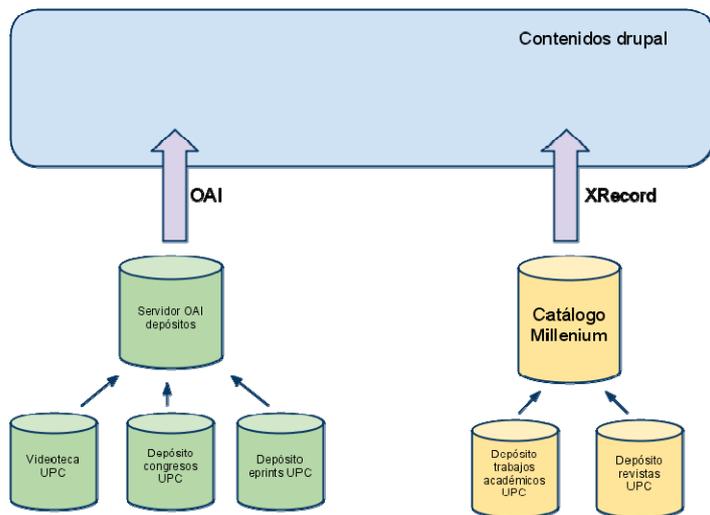


Figura 3. Estrategia seguida: recursos externos como los contenidos del OPAC y repositorios, se importaran como contenidos a DRUPAL

MÓDULOS DE IMPORTACIÓN

El primer motor de importación resultó sencillo de implementar, para ello hay que fijarse en un módulo denominado Millennium (<http://drupal.org/project/millennium>). Entre sus características principales destacan mostrar en tiempo real la disponibilidad de ese elemento en la biblioteca, mapear los diferentes campos del registro MARC a taxonomías de DRUPAL creadas previamente, mostrar la cubierta del libro configurando un servicio de imágenes de carátulas tipo Amazon, la posibilidad de ver una previa del libro si este se encuentra en “Google Books,” así como poder reservar el libro a través de una pasarela con el catálogo.

También permite tener un control sobre de que forma se deben importar los registros de Millennium a DRUPAL, si de forma manual o automática, siendo esta última la mas interesante, ya que el sistema es capaz de detectar cuando ha llegado al final de todos los elementos a importar y empezar de nuevo la reimportación para mantener sincronizados todos los datos que provienen del catálogo.

Para el segundo motor de importación existen diferentes proyectos en marcha en la comunidad DRUPAL, pero ninguno de ellos tiene aun publicado el código de manera oficial, o bien aquellos que lo tienen publicado no se acaban de adaptar a lo que se necesita. Algunos de estos proyectos son Extensible Catalog (<http://www.extensiblecatalog.org/>), OAI-PMH Module (<http://drupal.org/project/oai2>) o DRUPAL OAI-MPH (http://drupal.org/project/oai_pmh).

El primero de estos proyectos tiene publicada una versión preliminar del código, no obstante, la aproximación a la solución difiere un poco de lo que se necesita, siendo el apartado en el que se mapean los metadatos a las taxonomías aquello que mas se encuentra a faltar. El segundo tiene desarrollado un servidor de metadatos OAI, pero no así un recolector. Es el tercero de ellos el que se aproximaba a lo que necesitábamos, aun así el código ni tan siquiera ha llegado a publicarse de forma oficial.

A pesar de ello desarrollamos una serie de funcionalidades sobre este código en fase embrionaria para dotarlo de funcionalidades parecidas a las del módulo de Millennium y que próximamente se liberará en la comunidad DRUPAL. Entre sus características principales destacan la de mapear los diferentes campos del registro OAI a taxonomías de DRUPAL creadas previamente, mostrar la cubierta del libro configurando un servicio de imágenes de carátulas tipo Amazon o la posibilidad de seleccionar que colecciones deben importarse a DRUPAL.

TAXONOMIAS

Llegado a este punto, disponíamos ya de todos los elementos necesarios dentro de DRUPAL para empezar a mostrar y/o agrupar los contenidos según las necesidades. En ese sentido la comunidad DRUPAL nos ofreció un módulo con una serie de interfaces para realizar una navegación de contenidos dirigida en base a las taxonomías. Esto es, ofrecer filtros de entrada, de forma que el usuario vaya acotando cada vez más los elementos que pueden resultar de su interés. Seleccionar Audiovisuales de Calatrava en inglés ahora es relativamente sencillo gracias al módulo Faceted Search (http://drupal.org/project/faceted_search)

Esta aplicación en si mismo no realiza nada, pero en combinación con otros módulos supone una herramienta muy potente para ayudar al usuario a

acotar los contenidos. No obstante existen otras aproximaciones al mismo problema con el añadido de disponer de un buscador integrado. Una de ellas es el módulo Apache Solr Search Integration (<http://drupal.org/project/apachesolr>), que es finalmente es seleccionado para implementarse.

El núcleo de este módulo es el buscador Lucene, ampliamente usado en la comunidad DSpace. Entre otras características el módulo ofrece acotar los resultados, de forma que después de una primera búsqueda, el usuario es capaz de refinarla en base a los términos de las taxonomías que mas aparecen en los resultados de la búsqueda. En la figura 4 podemos observar una pantalla ejemplo con resultados de búsqueda.

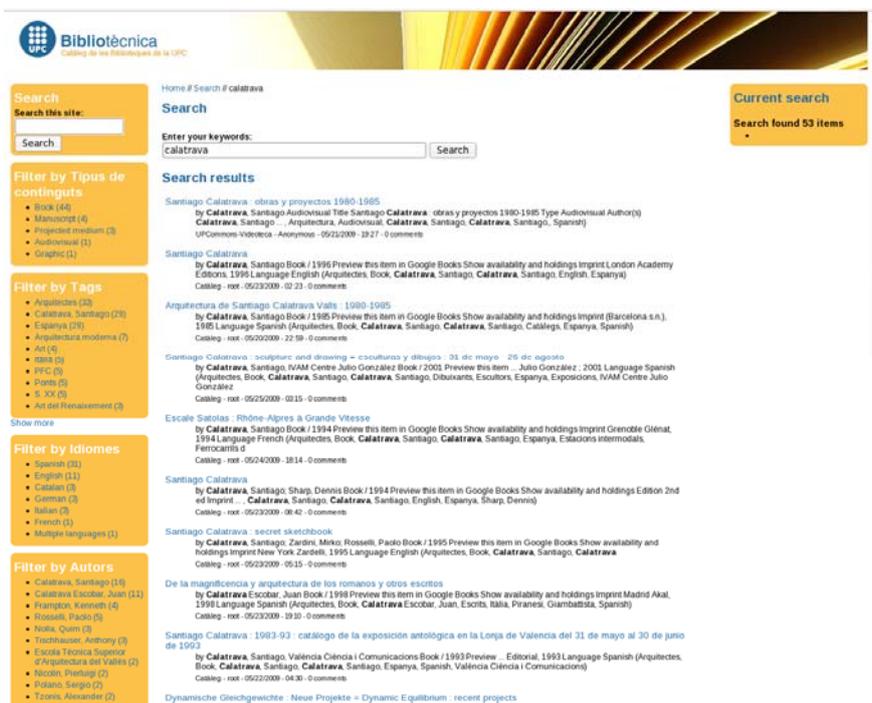


Figura 4. Resultados de la búsqueda realizada mediante el término "calatrava"

Otra característica que resulta interesante es la posibilidad de dar mas peso a unos tipos de contenidos sobre otros, excluir ciertos elementos de la búsqueda o la recomendación de otros elementos parecidos al que se está

visualizando. De esta forma no solo conseguimos que el usuario sea capaz de encontrar información relacionada con la temática buscada, si no que es capaz de filtrarla. En la figura 5 se puede ver la recomendación que realiza el sistema cuando se está consultando un resultado de la búsqueda.



Figura 5. El bloque "More like this" nos ofrece elementos interesantes parecidos al que se está visualizando

RENDIMIENTO

Un aspecto que nos inquietaba y que considerábamos un punto crítico era rendimiento. Si tenemos en cuenta que se está cargando aproximadamente 350.000 elementos dentro de DRUPAL y la existencia de taxonomías con unos 400.000 elementos. En este sentido, la comunidad DRUPAL, ofrece un módulo para la gestión de enormes taxonomías. Se trata de LeftandRight - Nested Set Taxonomy (<http://drupal.org/project/leftandright>) que es capaz de trabajar con taxonomías de más de 2.000.000 de elementos sin que para ello se penalice el rendimiento del sistema.

MÓDULOS PARA NUEVAS FUNCIONALIDADES

Resuelto el tema de la integración, nos planteamos en paralelo añadir nuevas funcionalidades y explotar los contenidos del site. Algunos de los módulos implementados han sido:

- Content Recommendation Engine (<http://drupal.org/project/cre>) que sirve para recomendar elementos de Drupal al más puro estilo de Amazon.

- User-to-user recommendation (<http://drupal.org/project/uurec>), utilizan filtrados colaborativos para recomendar elementos que pueden ser de interés para el usuario.
- Affinity pone sobre aviso a los usuarios que tienen intereses parecidos (<http://drupal.org/project/affinity>) en base a los elementos que se están consultando.

Existen otros módulos que sin ser herramientas para explotar los contenidos ayudaran en la gestión de la biblioteca. Como ejemplo cabe destacar Reservation module (<http://drupal.org/project/reservation>) que permite ofrecer un sistema de reservas dentro del site. Prestamos de portátiles, reserva de sala de estudio son sólo unos ejemplos.

CONCLUSIONES

Dentro del proceso de redefinición de la web de las bibliotecas de la UPC hacia un portal 2.0 los dos aspectos más relevantes eran por un lado seguir con el mantenimiento y gestión de forma descentralizada y ampliar esta funcionalidad a los propietarios de los procesos sin la necesidad de conocimientos HTML y/o la derivación de este mantenimiento a los especialistas “digitales”. El segundo aspecto, era como hacer partícipe al usuario, y como integrar en una misma interface las diferentes aplicaciones y servicios ofertados.

De forma tradicional, las bibliotecas han presentado en sus portales y de forma organizada, diferentes opciones para acceder a sus contenidos en función de su tipología y/o formato: OPAC, bases de datos, revistas digitales, repositorio, etc...

Como resultado de la encuesta de usuarios y del proceso de reflexión sobre cómo debería ser Bibliotecas 2.0, afloró en todos los casos dos aspectos críticos: por un lado el crecimiento y la dispersión de los contenidos hecho que obliga al usuario a realizar diversas búsquedas en diversas aplicaciones y en algunos casos con diferentes taxonomías e idiomas. El segundo aspecto era la complejidad en las búsquedas. Lo que nos

reclamaban los usuarios era una única interfaz para acceder a los diferentes contenidos.

Básicamente, el planteamiento y debate era este: “Si busco en la biblioteca UPC información sobre polímeros. A priori, ¿lo que necesito es un libro?, ¿o una revista digital?, ¿o un “preprint” del repositorio?, ¿o quizás el contenido de una base de datos?, ¿una tesis doctoral?.....” lo que hasta la fecha implica diferentes consultas. El reto fue plantearlo a revés: que escriba “polímeros” en una caja de búsqueda y en la lista de resultados le indicamos que tenemos y luego que el usuario decida y aportemos valor con alguna funcionalidad que además recomiende.

Entre las acciones para encontrar soluciones para la gestión y mantenimiento, así como para el usuario, la más relevante fue la detección de la necesidad de un CMS y el análisis de varios productos para detectar cual era el más pertinente para las bibliotecas de la UPC.

El resultado final, ha sido la adopción de DRUPAL (7) como CMS al que se han ido añadiendo módulos en función de los requerimientos de las bibliotecas y entre los que cabe destacar la integración de diferentes aplicaciones, especialmente el repositorio institucional (UPCommons) y el OPAC de Millenium y la funcionalidad de la búsqueda en una sola caja, con la posibilidad de “refinar” los resultados. Además de que la propia aplicación “recomienda” contenidos a partir de las taxonomías y que el usuario puede etiquetar, evaluar y/o opinar sobre los contenidos.

Como experiencia de su implementación, podemos afirmar que DRUPAL pone al alcance de todas las bibliotecas, herramientas, soluciones y técnicas que hasta ahora solo estaban al alcance de unas pocas empresas (8). Por un lado simplificamos el proceso de mantenimiento y gestión incluyendo la definición de flujos de trabajo adaptados a las necesidades de cada biblioteca UPC, y por otro integramos y simplificamos el acceso a los contenidos ofreciendo a los usuarios la posibilidad de participación.

El nuevo portal de las bibliotecas UPC: Bibliotecas 2.0 estará operativo al inicio de curso 2009-2010 (figura 6), Los próximos pasos son agregar nuevas funcionalidades especialmente orientadas a la segmentación de usuarios y a la personalización y intentar hallar una solución para la indexación, recuperación y naturalmente integración en DRUPAL, de los contenidos suscritos por medio de bases de datos o paquetes de publicaciones actualmente gestionado por Metalib

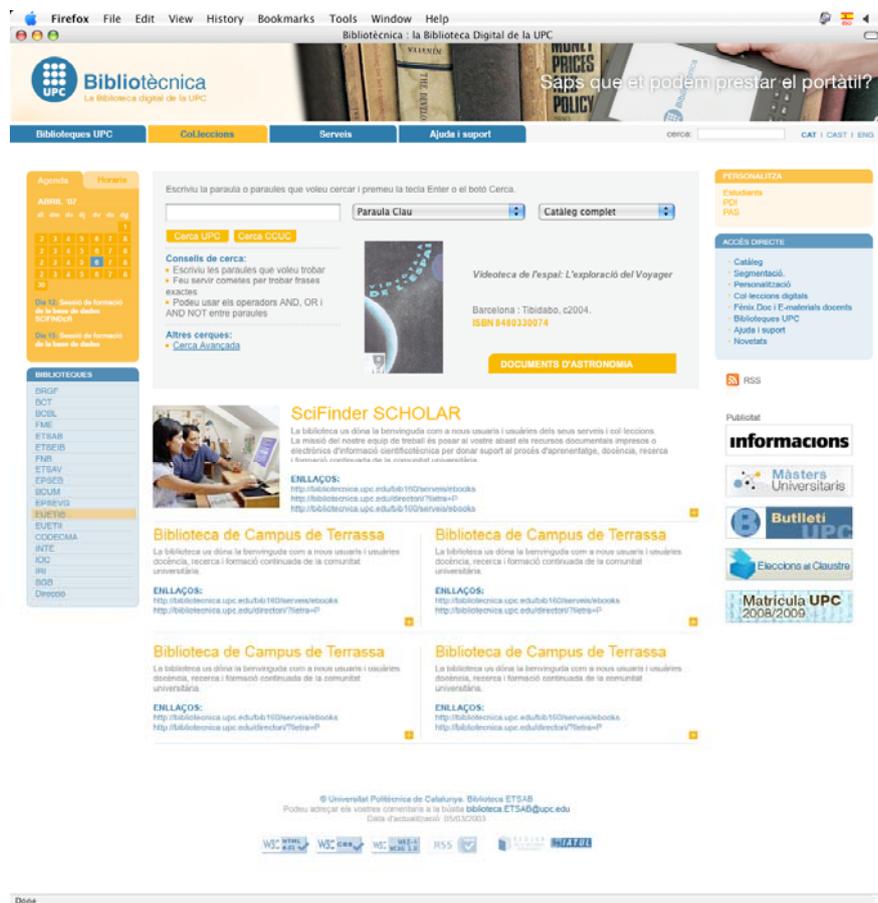


Figura 6. Aproximación al nuevo portal Bibliotecàtica 2.0

BIBLIOGRAFÍA

1. **Rodriguez-Gairin, Josep Manuel; Serrano-Muñoz, Jordi.** *El WEB com a mitjà d'estructuració i distribució d'informació bibliotecària [El WEB como medio de estructuración y distribución de información bibliotecaria]*. En: 5es. Jornades Catalanes de Documentació Barcelona : COBDC, SOCADI, 1995.. pp. 539-548.

http://bd.ub.es/pub/rzgairin/elweb_cs.htm [Consulta 14/06/09]

2. **Prats Prat, Jordi; Rico Millán, Oriol; Calderón Peñuela, Francisca; Majó, Rosa** *Bibliotecàtica: la Biblioteca Digital de la UPC. Descripció i valors afegits*. En: Jornades

Técnicas de Rediris. Palma de Mallorca : 2003. <http://hdl.handle.net/2117/2895> [Consulta 14/06/09]

3. **UPC. Servei de Biblioteques i Documentació APRÉN 2007-2010. 4t Pla estratègic Biblioteques UPC.** Barcelona: la Universidad, 2007.
http://biblioteca.upc.edu/Pla_estrategic/apren2010.pdf [Consulta 14/06/09]

4. **CMSMatrix** <http://www.cmsmatrix.org> [Consulta 14/06/09]

5. **Eden, Bradford Lee (ed.)** *Content management systems in libraries : case studies.* Lanham, Md. : Scarecrow Press, 2008

6. **O'Reilly, Tim.** *What Is Web 2.0.* <http://oreilly.com/web2/archive/what-is-web-20.html> [Consulta 14/06/09]

7. **Prats Prat, Jordi; Prieto Jiménez, Antonio Juan; Serrano-Muñoz, Jordi; Rico Millán, Oriol.** [Preprint] *Biblioteca Digital de la UPC. Drupal como gestión de contenidos 2.0* <http://hdl.handle.net/2117/2914> [Consulta 14/06/09]

8. *Drupalib A place for library drupalers to hang out* *drupalib A place for library drupalers to hang out* <http://drupalib.interoperating.info/> [Consulta 14/06/09]