



NEXUS: SISTEMA PARA FACILITAR LA DIFUSIÓN DE LA INFORMACIÓN EN BIBLIOTECAS UNIVERSITARIAS



José A. Senso, Amed Leiva-Mederos y Sandor Domínguez-Velasco



José A. Senso es doctor en documentación y profesor titular de universidad en el *Depto. de Información y Comunicación* de la *Univ. de Granada*, donde imparte docencia sobre alfabetización informacional en traducción e interpretación, y tecnologías de web semántica. Pertenece al grupo de investigación de lexicografía contrastiva *LexiCon*. Investiga en ontologías y gestión automatizada de corpus.

<http://orcid.org/0000-0002-6553-6522>

Univ. de Granada, Depto. de Información y Comunicación
Campus de Cartuja. 18071 Granada, España
jsenso@ugr.es



Amed Leiva-Mederos es doctor en documentación y profesor del *Departamento de Ciencias de la Información* de la *Facultad de Ciencias de la Información* de la *Universidad Central Marta Abreu de las Villas (UCLV)*. Dirige el proyecto de extracción de información en patentes de la *UCLV*. Ha sido ponente en diversos eventos nacionales e internacionales.

<http://orcid.org/0000-0002-9144-5018>

Universidad Central Marta Abreu de las Villas
Carr. Camajuaní, Km 5 ½. Santa Clara, 54930 Villa Clara, Cuba
amed@uclv.edu.cu



Sandor Domínguez-Velasco es ingeniero en control automático y sistemas computacionales por la *Universidad Central Marta Abreu de las Villas (UCLV)*. Profesor de programación del *Departamento de Automática* de la *UCLV*, ha impartido docencia en varios diplomados y curso de formación para el diseño de sistemas computacionales.

<http://orcid.org/0000-0002-1410-7445>

Universidad Central Marta Abreu de las Villas
Carr. Camajuaní, Km 5 ½. Santa Clara, 54930 Villa Clara, Cuba
sandore@uclv.edu.cu

Resumen

Como objetivo para lograr una mayor visibilidad de las aportaciones e investigaciones generadas por los miembros de la comunidad universitaria, se plantea el desarrollo de un cms (*content management system*) que trabaje junto a la iniciativa *Sioc (semantically-interlinked online communities)* gestionado dentro de la biblioteca universitaria. Esto permite la interoperatividad entre los sistemas de información de la propia biblioteca y la difusión de la información en redes sociales. Se presenta *Nexus*, una herramienta que facilita la interacción entre investigadores y profesores universitarios por medio de microformatos, empleando *Drupal* como cms. Se explican las características del servicio, sus nodos y módulos, y las prestaciones que ofrece bajo la filosofía de *linked data*.

Palabras clave

Bibliotecas universitarias, Servicios de información, *Linked data*, CMS, *Drupal*, Redes sociales.

Title: Nexus: a system to facilitate the dissemination of information in academic libraries

Abstract

With the objective of achieving greater visibility for the research and scholarly contributions generated by members of the university community, the development of a content management system (cms) is proposed that would work with the *Semantically-interlinked online communities (Sioc)* initiative, managed within the academic library. This enables interoperability between the information systems of the library and the dissemination of information in social networks. *Nexus* is a tool that facilitates interaction between researchers and academics using microformats and *Drupal* as a cms. This paper explains the features of the service, its nodes and modules, and the services offered under the philosophy of *linked data*.

Artículo recibido el 18-09-2012

Aceptación definitiva: 12-01-2013

Keywords

Academic libraries, Information services, Linked data, CMS, Drupal, Social networks.

Senso, José A.; Leiva-Mederos, Amed; Domínguez-Velasco, Sandor (2013). "Nexus: sistema para facilitar la difusión de la información en bibliotecas universitarias". *El profesional de la información*, marzo-abril, v. 22, n. 2, pp. 163-172.

<http://dx.doi.org/10.3145/epi.2013.mar.11>

1. Introducción

Las bibliotecas universitarias españolas, durante su proceso de transformación en crais (centros de recursos para el aprendizaje y la investigación) a imagen de los *Integrated learning centres* ingleses y los *Information commons* norteamericanos, han llevado a cabo cambios con el fin de hacer más accesible sus recursos a los usuarios y ampliar así la oferta con productos más ajustados a sus objetivos: la docencia y la investigación. Las bibliotecas universitarias han adoptado un papel más activo en la gestión de la información, tratando de anticiparse a las necesidades de sus usuarios, ofreciéndoles información casi personalizada adecuada a cada perfil. Para ello han tenido que emplear diversas herramientas (**Amante; Extremeño, 2012**).

Los crais son una plataforma única para el intercambio de información dentro de un ámbito de cooperación tanto interuniversitario como investigador (**Zamora, 2012**). En este afán por aglutinar lo mejor posible la inmensa cantidad de información de que disponen, no han terminado de resolver (salvo honrosas excepciones) de manera satisfactoria un problema de calado: lograr una mayor visibilidad de las aportaciones e investigaciones generadas por los miembros de su comunidad universitaria.

En la mayoría de los casos, hay dos opciones. Por un lado, integrar la producción científica de la institución en el catálogo de la universidad, mezclando las referencias bibliográficas (y, en ocasiones, el texto completo) a investigaciones y sus productos, con el resto del fondo bibliotecario. De esa manera, y por medio de la interfaz común del opac, la biblioteca integra contenidos propios y ajenos, eliminando la identidad y el sentido local que tienen los primeros. La otra opción se ha centrado en la creación de repositorios institucionales, lo que facilita que los contenidos sean recolectados por robots OAI-PMH promoviendo su mayor difusión.

Esto último no siempre es así, y son numerosos los casos de repositorios que no ofrecen sus datos a los recolectores, al no tener actualizada correctamente la interfaz OAI-PMH (**Barrueco-Cruz; García-Testal, 2009**). Otro de los problemas que se suele dar con frecuencia es la falta de registro en directorios nacionales o internacionales (**Poornima, 2012**). Pero lo que parece ser una constante es la pobreza en la descripción de los objetos incluidos, empleando sólo el sistema de metadatos OAI-DC (**Schöpfel; Bescond; Probst, 2012; Chuttur, 2012; Melero et al., 2009**), por lo que no se logra una interoperabilidad total entre repositorios (**Hebrang-Grgić, 2011**). El problema es, pues, que la información almacenada en el repositorio sólo forma parte de la internet invisible, al ser poco accesible por los servicios de búsqueda en la Web (**Merlino-Santesteban, 2012**).

Además, a pesar de contar con opacs y repositorios, los investigadores acuden a portales como *Academia*, *ResearchGate* o *ImResearch* para dar a conocer sus aportaciones (**Merlo-Vega et al., 2011**), posiblemente por la necesidad de formar redes de trabajo o, simplemente, aumentar el impacto y visibilidad de sus aportaciones, es decir, lograr más citas.

Actualmente no hay más opciones viables para prestar este servicio dentro de las plataformas con las que la biblioteca trabaja convencionalmente. Sin embargo, existe una vía, poco explorada, y que presenta múltiples ventajas: el uso de cms (*content management systems*) que trabajan junto a la iniciativa sioc (*semantically-interlinked online communities*), permite generar modelos de organización de información capaces de mezclar diversos tipos de microformatos y facilitar así la interoperatividad entre sistemas de información y redes sociales. Los microformatos son un mecanismo sencillo para agregar significado semántico a un documento que, sin dejar de ser legible para las personas, aporta estructura al sistema de información de modo que es posible procesar su código fuente (**Méndez; Bravo; López, 2007**).

Las bibliotecas universitarias no han logrado aportar más visibilidad a las investigaciones generadas por los miembros de su comunidad universitaria

Sobre esta premisa se construye este artículo, que describe el proceso de creación de *Nexus*, una herramienta que facilita la interacción entre investigadores y profesores universitarios por medio de microformatos, empleando *Drupal* como cms. Es una plataforma de trabajo en entorno web capaz de manejar grandes conjuntos de datos transformados en descripciones rdf (*resource description framework*), y que tiene como principal objetivo la difusión del conocimiento científico a través de las bibliotecas académicas.

2. Nexus

Aunque en la actualidad existe gran cantidad de cms de código abierto, son pocos los capaces de combinar los elementos de la web semántica en la gestión de los contenidos.

Drupal proporciona una solución genérica para la publicación web de contenidos creados por personas que actúan de manera individual o dentro de organizaciones y comunidades de cualquier tamaño. Todos los cms emplean bases de datos relacionales para almacenar la información (*MySQL, PostgreSQL, SQLite, SQL Server...*). Sin embargo, trabajar con información semántica requiere del uso de un modelo

de descripción rdf, que difiere de esos modelos de bases de datos (Allemand; Hendler, 2011). Para que el cms pueda operar con este tipo de datos es necesario emplear una capa intermedia que permita almacenar datos rdf en tablas relacionales (Pastor-Sánchez, 2012). Ahí reside la principal diferencia entre los cms tradicionales y los semánticos, y no tanto en el hecho de que existan otras opciones como rdfs o la posibilidad de trabajar con microformatos.

Los cms semánticos facilitan la gestión y desarrollo de software, la colaboración entre investigadores y la publicación de documentos, además de operar con *linked data*. Son eficientes en el manejo de grandes conjuntos de datos distribuidos, la visualización de datos y la construcción de herramientas de análisis (Mooney; Baenziger, 2007).

Los cms tradicionales —en contraposición a los semánticos— presentan varios problemas a la hora de gestionar la información:

- Los datos que presentan no sirven para reflejar los flujos de trabajo y la totalidad de las prestaciones que se demandan.
- Poseen poca relación entre el contenido que muestran (texto, imágenes, vídeo, datos, etc.) y el contexto donde se ofrece dicho contenido.
- Las soluciones informáticas que ofrecen no reflejan las necesidades de los usuarios.

Desde la versión 7 *Drupal* incluye, de forma nativa, soporte a rdfs (conjunto de extensiones de xhtml propuestas por el consorcio W3C para introducir semántica en los documentos), lo que facilita el posicionamiento de las páginas web creadas con este sistema, además de promover la interoperatividad de datos. Por todos estos motivos se escogió este software para la creación de *Nexus*.

2.1. Drupal

Es un gestor de contenidos creado en 1999 por Dries Buytaert y desarrollado bajo licencia GNU dos años más tarde. Crear un sitio web en *Drupal* consiste en combinar varios “bloques” con el fin de adaptar la funcionalidad del sitio se-

gún las necesidades específicas. Además, es un marco de gestión de contenidos (*content management framework*) (Byron; Berry; De-Bondt, 2012). La información se almacena en una base de datos relacional (puede trabajar con *MySQL*, *PostgreSQL*, *SQLite...*) por medio del lenguaje de programación php (Pastor-Sánchez, 2011).

Permite la publicación de datos en formato rdf, además de soportar otros formatos como n-triples, json, xml, rss 1.0 y turtle. Gestiona las uris de los datos rdf publicados y proporciona la administración de un *endpoint sparql* para la consulta de los datos. Permite la personalización de los campos rdf y *namespaces* (Alonso-Sierra; Ortiz-Muñoz; Hidalgo-Delgado, 2012).

Los investigadores acuden a portales como *Academia*, *ResearchGate*, *ImResearch* o *Cosis.net* para obtener servicios que no les ofrecen sus bibliotecas

Es un software muy flexible, con un mecanismo de seguridad robusto, y suficiente documentación online. *Drupal* dispone de módulos para agregar capacidades adicionales a medida que los sitios web se van desarrollando (Lillevig; Stevens, 2009).

2.2. Consideraciones del diseño de Nexus

Con el fin de poder gestionar mejor los datos introducidos en el sistema se optó por emplear microformatos, entre los que se encuentran skos (*simple knowledge organization system*), Dublin core y foaf (*friend of a friend*). Así se puede construir un sistema escalable que facilita la conexión entre los usuarios de la comunidad académica con los contenidos que ésta genera. El diseño tuvo en cuenta los siguientes elementos:

- gran número de profesores e investigadores de universidades que son parte activa y pasiva de la Red;
- necesidades de los investigadores y docentes como usuarios de información;
- riqueza epistémica y lingüística del dominio en que se implementa el producto;
- flexibilidad de los microformatos y posibilidades de su modificación;
- facilidades de *linked data*;
- capacidad semántica de la plataforma;
- exportación de los contenidos y las citas de los trabajos a gestores bibliográficos (*Procite*, *EndNote...*);

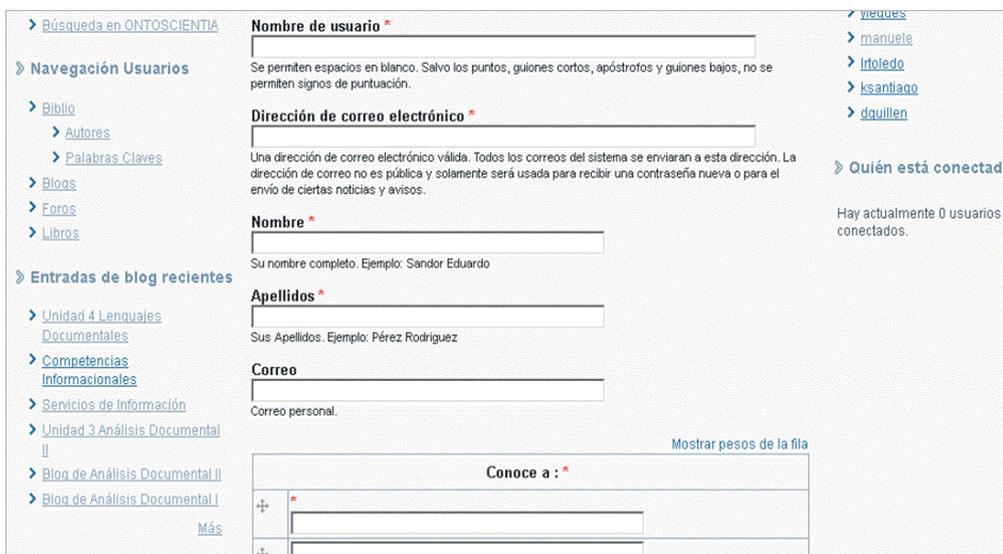


Figura 1. Aspecto de la interfaz para introducir datos personales. Los datos de cada campo están asociados al microformato foaf correspondiente

- existencia de una licencia de *open access* que regula el contenido y el uso de la información en el sistema.

Para hacer frente a estos requisitos de diseño y los desafíos que desde el punto de vista de la gestión de la información supone una aplicación de este tipo, se optó por ajustarse exclusivamente a las necesidades de los usuarios a los que se dedica, tratando de evitar las posibles barreras en el diseño a la hora de construir, compartir y administrar datos. También se intentó adaptar el producto final a los requerimientos de cada disciplina académica, estableciendo colaboraciones con líderes científicos, con la idea de que sirvieran de apoyo en la elaboración del software.

2.3. Servicio de información de Nexus

Se utilizó la metodología planteada por Garrido y Tramullas (2006) por ser la que mejor muestra los aspectos a tener en cuenta para el correcto diseño de ofertas de información digital.

Según esta metodología hay cuatro aspectos a considerar antes de diseñar e implementar cualquier servicio de información digital:

- elementos y entorno;
- plan de gestión de los recursos de información;
- modelo conceptual;
- estudio de usuarios.

Por las características particulares de Nexus, sólo se usaron el primero y el último de esos aspectos. El estudio de necesidades de información consistió en determinar los principales frentes de investigación de los usuarios potenciales, los temas más demandados, los horarios de servicio, las formas de emitir los resultados o formatos de entrega de información, así como los roles que tendrían los usuarios en el sistema para su administración, mantenimiento y uso.

Una vez hecho esto se realizaron las actividades para la creación del servicio:

- Estudiar las necesidades de los usuarios potenciales del servicio.
- Desarrollar los procesos de ingeniería de software.
- Desarrollar autómatas y programar modelos de tratamiento de texto.
- Prueba del programa.
- Entrenamiento del personal y confección de la documentación del software.
- Fase piloto del software.
- Publicación y divulgación del servicio.

3. Arquitectura y flujo de trabajo

La arquitectura de Nexus se basa esencialmente en las potencialidades de las ontologías, las facilidades que brindan los microformatos y las bondades que *linked data* ofrece para el trabajo con datos remotos.

Esta plataforma emplea, fundamentalmente, cuatro estándares:

Foaf

Permite reunir datos de personas. Es un léxico de un vocabulario rdf para delimitar semánticamente información per-

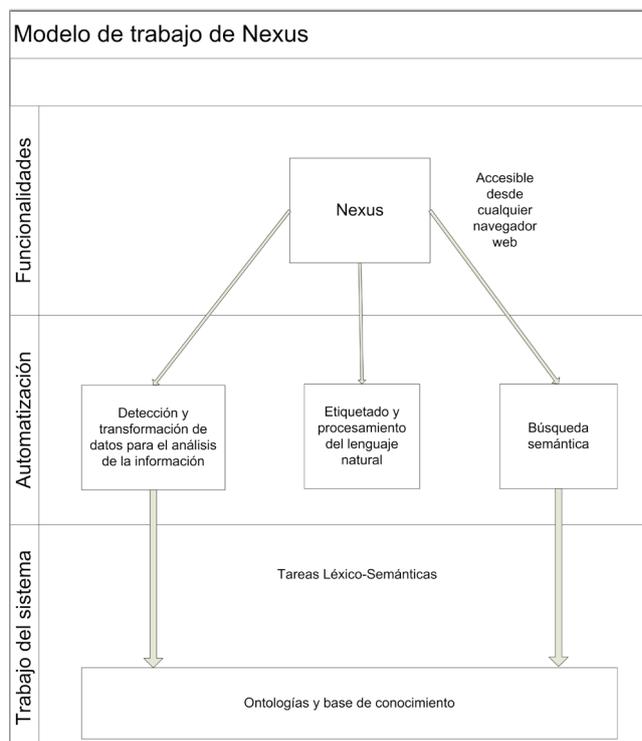


Figura 2. Cuando un usuario entra por primera vez en la plataforma, sus datos personales se almacenan en foaf, sus documentos y publicaciones se describen por medio de microformatos de Dublin core, sus temas de investigación se equiparan a clases en skos y, por último, sioc une las ontologías resultantes de cada especialidad, de modo que no haya solapamiento en las funciones y todos los datos puedan ser utilizados para la construcción de redes sociales.

sonal muy extendida en redes sociales como Facebook. Las clases que genera son eficientes y pueden ser enriquecidas con el uso de otros formatos, como Dublin core. En este proyecto el usuario se encuentra con un formulario que facilita la recogida de datos: correo electrónico, personas a las que conoce, subir un archivo con su currículum personal..., y toda esta información se almacena en la tabla *Person* de la base de datos de Drupal que, además, asocia cada uno de estos campos a las etiquetas foaf correspondientes. Con el fin de aprovecharlo mejor se ha implementado un módulo específico que facilita la localización de investigadores (que ya sean usuarios del sistema).

Sioc

Se emplea el módulo facilitado por Drupal. Facilita la descripción de comunidades o grupos dentro de un cms. Hace que diversos usuarios, mediante foros de discusión online o blogs, puedan asumir diversos roles de trabajo para intercambiar conocimiento. Con sioc las comunidades de usuarios online se convierten en nodos que se vinculan entre sí, y a los tipos de contenido. Gracias a sus capacidades ontológicas no sólo es posible vincular comunidades de usuarios, sino aprovechar los diversos predicados de los datos rdf para mezclarlos y reutilizarlos sin que provoquen problemas en el sistema (Breslin et al., 2006). Sioc dentro de Drupal facilita la creación de metadatos en los heterogéneos tipos de contenidos que maneja este cms como: blogs, nodos, comentarios, foros, usuarios y roles, lo que permite que los diversos mensajes registrados por autor o tema sean atribuidos a cada rol.

Dublin core

Se han creado cuatro tipos de contenidos, que se añaden a los que ya incluye *Drupal*: resumen, imagen, autoridad y artículo. Cada uno posee *namespaces* específicos para tratar la información de Dublin core y, con ello, se facilita la recuperación de información por medio de sparql. A cada tipo de contenido se le ha añadido un conjunto diferente de etiquetas, para no generar redundancias en la base de datos:

- resumen cuenta con *dc:title*, *dc:creator*, *dc:contributor* y *dc:date*;
- imagen emplea *dc:title*, *dc:subject* y *dc:resource*;
- autoridad usa *dc:creator* y *dc:contributor* junto a otra serie de propiedades foaf;
- artículo emplea *dc:subject*, *dc:title* y *dc:creator*.

Los sistemas de información creados sobre cms con soporte sioc permiten modelos de organización de información capaces de integrar microformatos y facilitar así la interoperatividad

Skos

Se utiliza en las taxonomías, vocabularios y en formalizaciones de tesauros en el entorno de la web semántica. Se emplea integrado dentro del módulo de taxonomía de *Drupal*.

El modelo funcional de *Nexus* (figura 2) tiene como premisa que el software facilite la automatización de procesos —especialmente los de comunicación dentro de la biblioteca— a partir del uso de algoritmos que posibiliten la detección de datos y el etiquetado semántico de la información.

Con el fin de facilitar la visualización de la producción del investigador/profesor, en el momento del registro *Nexus* activa una api de *Google* que conecta los datos que se introducen en el sistema con *Google scholar*. Al mismo tiempo, el programa le asigna automáticamente un rol (de acuerdo con las características de su CV) y, con ello, un espacio para su blog personal (para publicitar sus investigaciones) o la posibilidad de colaborar con el blog de su grupo de investigación. Dicho blog puede integrar contenidos tanto propios (artículos de miembros del grupo publicados en revistas electrónicas) como de otras bases de datos tanto (*Facebook*, *Google scholar*, *Wikipedia*...) o suscritas por la biblioteca, y todo ello por medio de una serie de ontologías que facilitan la gestión de los términos empleados, el etiquetado semántico automático y la búsqueda de recursos (figura 3).

3.1. Nodos y módulos

En *Drupal* los nodos son la unidad de información básica en la que se almacenan los contenidos. Un nodo es una abstracción que se corresponde con una estructura de tablas interrelacionadas de la base de datos del sitio web.

Todo el contenido de *Nexus* se almacena en nodos que se pueden agrupar de 6 formas. La principal, creada con sioc, es la que une los lenguajes formales y los microformatos, y tiene como tablas asociadas a *Node* (nombre de la tabla donde se almacena la información de los nodos en *Drupal*, y que controla todas las prestaciones de la red social), *User* (administra los datos de los usuarios) y *Rol* (asigna permisos). Los elementos de foaf se conectan a *Rol* y a *User* para que los usuarios que se registren en el sistema tengan un rol y a la vez posean datos personales suficientes para incluirse en la red social que se conforma con todos los usuarios. Los elementos de skos (empleados para indizar temas de investigación) se unen a la tabla *Node*, de esa forma se enlazan con los metadatos Dublin core y foaf.

Los módulos de *Drupal* permiten añadir nuevas funciones para adaptar el cms a las necesidades de cada sitio. *Nexus* no prescinde de módulos esenciales de *Drupal*, de hecho, los módulos que se han creado nuevos dependen de éstos. La mayoría de los módulos que se explican en el siguiente apartado dependen de *Taxonomy*, *internacionalización (i18n)* y sus variantes, *Taxonomy_nco*, *Biblio* y otros que se explican a lo largo de este trabajo. Pero no necesariamente se emplean al completo, sino complementando módulos propios. Así, por ejemplo, en *Nexus* se han adaptado los

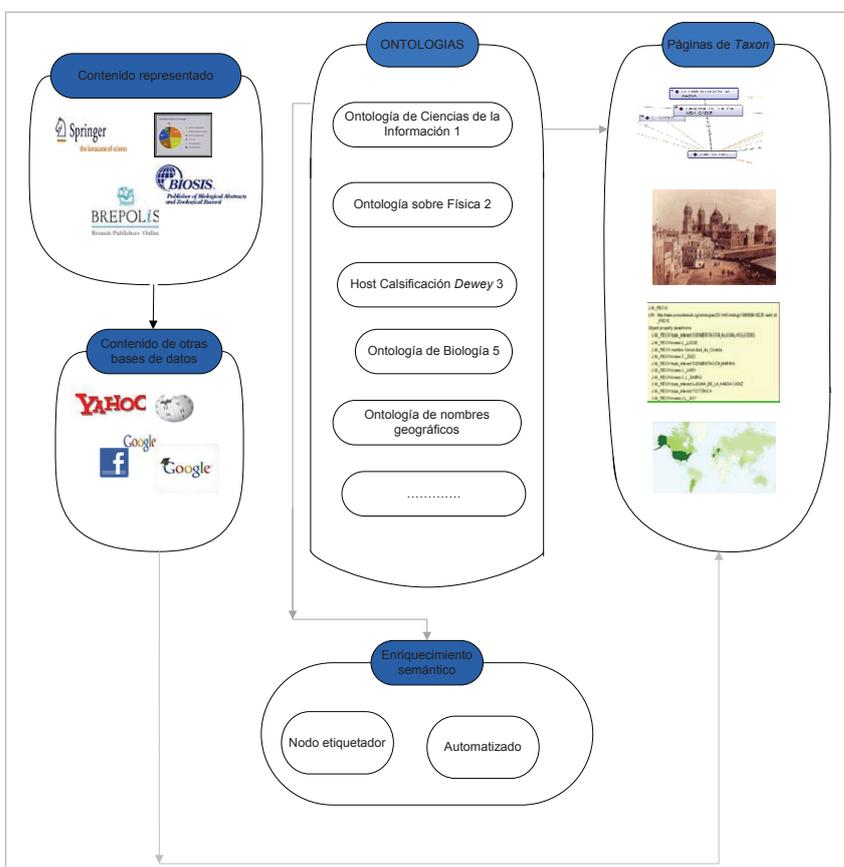


Figura 3. Gestión de los contenidos

Nombre	Versión	Uso	URL
<i>Nexus-users</i>	Candidata	Gestión de usuarios	http://code.google.com/p/nexus-users
<i>Skosdeweycdu</i>	Beta	Gestor de vocabularios Skos	http://code.google.com/p/nexus-skosdeweycdu
<i>Imagen.tools</i>	Definitiva	Mejora en la carga y gestión de imágenes	http://code.google.com/p/nexus-image-tools
<i>Nexus-Puertotex</i>	Beta	Creación de resúmenes automáticos	http://code.google.com/p/nexus-puertotex
<i>Nexus-Analyzer</i>	Definitiva	Análisis de la información	http://code.google.com/p/nexus-analyzer
<i>Nexus-control autoridades</i>	Alpha	Gestión de las autoridades	http://code.google.com/p/nexus-control-autoridades
<i>Nexus-Onto y Nexus-subject</i>	Alpha	Mejoras en el sistema de búsqueda	http://code.google.com/p/nexus-onto http://code.google.com/p/nexus-subject

Tabla 1. Relación de módulos creados específicamente para *Nexus*. Todos se encuentran disponibles en *Google code* bajo licencia *GNU general public license* v. 3.0 y gestionados con el control de versiones *Git*

módulos *Taxonomy* y *Autotag* para crear un anotador semántico capaz de etiquetar el contenido introducido en el sistema sin ceñirse a un léxico particular. Para ello se une la taxonomía descrita para cada especialidad (en skos) con anotaciones individuales de los usuarios. Para la gestión del vocabulario skos se creó un módulo específico, denominado *Skosdeweycdu*, que equipara términos de la taxonomía con notaciones de las clasificaciones Dewey y CDU. La idea ha sido permitir elaborar una taxonomía de manera similar a como se realiza con la api de *Agrovoc* (Sánchez, 2011). Estos lenguajes, a su vez, están equiparados con los *LCSH* (*Library of Congress subject headings*).

Los datos con los que trabaja *Skosdeweycdu* de *LCSH* se han obtenido en formato skos-rdf/xml y se han ubicado de acuerdo con la estructura taxonómica que recomienda la *Library of Congress* (2011): encabezamientos de materia general, subdivisiones geográficas directas, subdivisiones geográficas indirectas, subdivisiones, subdivisiones temáticas, subdivisiones de género-forma, subdivisiones temporales y subdivisiones lingüísticas. Las clases, subclases e instancias obtenidas de Dewey y CDU se depositan de forma automática en el campo sinónimos de la taxonomía. Para ello se ha empleado la aplicación *Colector* (Domínguez-Velasco, 2010), que carga los términos de los lenguajes de clasificación en un archivo en formato rtf para, después, insertar el *namespace* de skos y la etiqueta *skos:altlabel*. De esta manera es más sencillo depurar los términos y llevarlos a la taxonomía.

<http://id.loc.gov/authorities/subjects.skos.rdf>

Lo que se consigue con este módulo es que el usuario pueda crear sus propios vocabularios o folksonomías, y que éstos se puedan vincular con términos normalizados, facilitando de esa manera la equiparación de términos propios con los almacenados en las bases de datos de la biblioteca universitaria. Esta opción ayuda a los investigadores/profesores a asociar el contenido temático de sus trabajos al mundo de su especialidad de acuerdo con sus competencias lingüísticas.

La gestión y visualización de referencias bibliográficas se lleva a cabo por medio del módulo *Biblio*. También se han empleado otros módulos básicos para gestionar los foros y los grupos.

3.2. Módulos especiales

La mayoría de las herramientas comentadas se encuentran

ya prediseñadas dentro de *Drupal*. Sin embargo con *Nexus* fue necesaria la creación de módulos específicos para que la plataforma pudiera ofrecer las prestaciones que los investigadores/profesores necesitan, y suponen el elemento diferenciador con respecto a otras aplicaciones similares. Además del citado módulo *Skosdeweycdu*, estos módulos (tabla 1) han sido:

- **Gestión de imágenes**

Aunque en *Drupal* existen varios módulos para la gestión y tratamiento de imágenes, la mayoría de ellos sólo permite presentar galerías, descargar imágenes y su edición. A partir del módulo *Image* se creó un segundo módulo, denominado *Imagen.tools*, capaz de mejorar la carga y visualización de imágenes –incluso si se encuentran dentro de un documento– y que, además, permite introducir anotaciones en las imágenes para que puedan ser consultadas a través del sistema de búsqueda global. Las imágenes que se suben a *Nexus* pueden estar asociadas a publicaciones que hayan sido cargadas en el módulo bibliográfico, por tanto se da a los usuarios la opción de gestionar las referencias del documento en que se encuentra la imagen que se está anotando.

- **Resumidor**

Los artículos científicos subidos a la plataforma son gestionados por este módulo para obtener resúmenes automáticos. Todo artículo científico a texto completo que se sube a *Nexus* se etiqueta de acuerdo con una dtd creada específicamente para ello. Las etiquetas del texto definen las partes esenciales del artículo científico (introducción, metodología, resultados y conclusiones) así como las oraciones y sus relaciones retóricas y semánticas. Un sistema de reglas de discurso identifica las oraciones de mayor carga semántica y selecciona los candidatos para construir el resumen. Una vez etiquetado el documento, sobre él actúa un agente inteligente capaz de utilizar las reglas de discurso para resumir el texto científico. También es posible resumir artículos científicos online, para que luego el usuario decida si consulta los documentos originales en Red. Todo este proceso sigue el modelo de trabajo definido en *PuertoTex* (Leiva-Mederos; Domínguez-Velasco; Senso, 2012).

- **Análisis de la información**

Las referencias bibliográficas de los artículos científicos recuperados online o subidos a la plataforma se extraen de forma automática para construir una matriz cuadrada que permite estudiar varios indicadores del análisis de un domi-

nio específico a partir de grafos generados en *svg* (*scalable vector graphics*). Los indicadores que se analizan con este módulo son redes de autores más productivos, editoriales, países, universidades, centros de investigación y temáticas.

• **Control de autoridades**

A *Nexus* también se le suma un sistema de control de autoridades basado en *viaf* (*virtual international authority file*). Para ello se utiliza un conversor de archivos marc a *rdf* utilizando las normas de *Oclc* e *Ifla* al respecto. Esto permite combinar la información almacenada en los catálogos de diversas bibliotecas universitarias y facilitar consultas masivas. Cuando se introduce un documento nuevo y el sistema detecta que existe una posible autoría, muestra todas las variantes de encabezamientos de autores que existen en bibliotecas universitarias previamente seleccionadas, y normaliza las entradas. Una vez hecho esto, y descrito por completo el documento, *Nexus* permite que cualquier biblioteca incorpore a sus catálogos estos registros mediante *Z39.50*. También es posible dar de alta nuevas autoridades de forma manual.

• **Búsqueda de información**

Una de las fortalezas de esta plataforma es su sistema de búsqueda (**Domínguez-Velasco; Senso; Leiva-Mederos, 2011**), creado específicamente de acuerdo con las especificaciones de *PuertoTex* (**Leiva-Mederos; Domínguez-Velasco; Senso, 2012**). Se emplea un algoritmo basado en grafos para construir búsquedas múltiples, sustentadas en un agente inteligente que emplea los términos de *skos* y los convierte en herramientas de búsqueda conceptual. Este tipo de búsqueda semántica no sólo usa la información almacenada en la plataforma, sino que también puede consultar los catálogos de otras bibliotecas universitarias. La búsqueda semántica se concreta gracias a la capacidad de su agente de búsqueda de cargar, de forma autónoma, información para los usuarios. Se encarga de buscar información empleando las materias que se encuentran en la taxonomía del sistema y recupera la información temática que ha sido formalizada de acuerdo al *dataset* de la *Library of Congress* que ha sido cargado con anterioridad.

<http://id.loc.gov/authorities/subjects>

Mediante los elementos que provee este *dataset* es posible ordenar diversos datos para hacer búsquedas específicas con el fin de que el agente localice información a través de la siguiente triplete: "Uri(s): Instancias, Esquema al que pertenece". Al recuperar información se devuelven al usuario diversos tipos de datos, como la colección temática que está asociada al término, o los conceptos más amplios y las fuentes donde se encuentra el recurso para que decida si

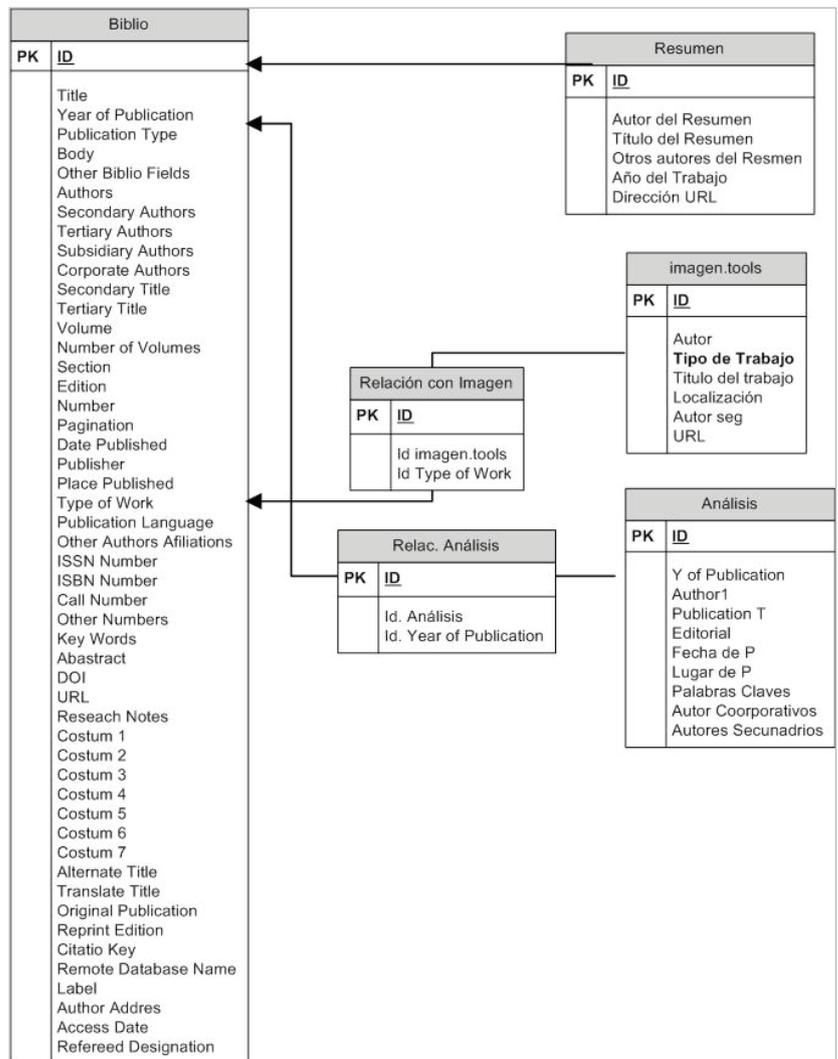


Figura 4. Se ha aprovechado la tabla del módulo *Biblio* para que sirva de nexo de unión entre varios de los módulos creados en este proyecto. De esta manera se logra una mayor coherencia en la información. Los datos introducidos en *Biblio* se procesan con el módulo *Nexus-analyzer* para el análisis del dominio. Las imágenes subidas se gestionan con el módulo *Image.tools* para facilitar su anotación. El resumen incluido en *Biblio* se genera automáticamente con el módulo *Puertotex*

consulta o no la información. El agente en cuestión monitorea toda la información nueva que hay sobre los temas que tiene en la taxonomía y la almacena en los perfiles de los usuarios.

Otra de las posibilidades que ofrece este agente de búsqueda es localizar información dentro de los registros de autoridad de las bibliotecas integrantes del proyecto *viaf*, las cuales proveen de información bibliográfica y datos de autoridad a las universidades y a las bibliotecas académicas. Desde *Nexus* se pueden explorar, visualizar y consultar datos de bases de datos externas. Para ello, se basa en el paradigma de la interoperabilidad gracias a la existencia de vocabularios especializados que permiten gestionar datos *rdf*.

Otro aspecto que se ha tenido en cuenta para la búsqueda externa de información es usar el núcleo central de *linked data*, la *DBpedia*, ya que a partir de ahí es posible enlazar con este tipo de proyectos (**Phipps; Hillman, 2011**). Para generar una consulta *linked data* con *Nexus* se emplean los *datasets* de los proyectos que se corresponden con las necesidades temáticas de los usuarios. El sistema registra

todos los *endpoints* de los *datasets* que pudieran ser susceptibles de ser empleados por los usuarios, en función a la temática a la que fueron asociados en el momento de darse de alta en el sistema. Cada *dataset* es descrito especificando los campos que han de mostrarse para realizar las consultas y declarando para cada campo el predicado rdf con sus correspondientes valores. Todos estos datos, ya formalizados, integran un tipo de contenido que es creado con el uri donde se encuentre físicamente el registro, lo que obliga a generar una vista para visualizar los datos, que se ubicará en forma de entrada de blog dentro de las páginas del sistema. Esta búsqueda se complementa con el módulo *Subject*, que permite realizar búsquedas por materias dentro de *Nexus*.
<http://code.google.com/p/nexus-onto>
<http://code.google.com/p/nexus-subject>

« *Nexus* conecta de forma automática la producción de un investigador con *Google Scholar*, con las bases de datos de la biblioteca y, además forma una red social »

4. Desarrollo actual y futuro

Tanto el marco global de trabajo (*Nexus framework*) como sus respectivos módulos se encuentran en fase beta desde julio de 2012. El objetivo es que para comienzos de 2013 se realice una versión candidata a definitiva, y que exista una versión de disponibilidad general a finales de 2013. En la actualidad hay una versión instalada en la *Universidad Central Marta Abreu* de Cuba en fase experimental, ya que es allí donde los autores han encontrado la masa crítica de usuarios suficiente como para poder realizar una evaluación exhaustiva del software. Además, su biblioteca universitaria presenta una situación similar a la expuesta en la introducción de este texto, con las necesidades y los problemas citados. *Nexus* aspira a desarrollarse como una plataforma *open source* y, aunque su origen se encuentra en ofrecer un servicio concreto al mundo académico, el objetivo final es que pueda ser empleado por diferentes grupos de usuarios. Para ello, en las próximas versiones se desarrollarán nuevos módulos para el tratamiento de diferentes tipos de información entre los que se encuentra el video y el sonido.

Desde el punto de vista técnico se plantea la posibilidad de utilizar algoritmos de agrupamiento que completarán el de resumen, ya que darán al sistema la posibilidad de aglutinar documentos según su contenido.

Los autores son conscientes de que *Drupal* no permite plantear mayores metas para la plataforma, principalmente debido a que las posibilidades de sus apis para rdf no están desarrolladas en su totalidad. Además el entorno de trabajo con sparql es aún incompleto, al igual que las opciones de

serialización y configuración del soporte rdf. Sin embargo *Drupal* ha demostrado ser capaz de soportar servicios de información con la dinámica de gestión de información que se requiere en la actualidad.

5. Conclusiones

Tanto investigadores como profesores universitarios necesitan aumentar la visibilidad de su producción. Hasta ahora esto lo logran peregrinando por las redes sociales especializadas, rellendo continuamente formularios de alta, actualizando perfiles de usuario, introduciendo el mismo artículo en diferentes plataformas una y otra vez. Teniendo en cuenta que esa visibilidad repercute no sólo en su beneficio sino también de la institución a la que pertenecen y de la biblioteca, es necesario que puedan contar con un programa que facilite este proceso. Si, además, tal programa está conectado con las bases de datos de la organización se logrará que la información fluya mejor entre los interesados.

El peso de estos programas, fórmulas o vías para lograr ese objetivo no tiene por qué recaer necesariamente sobre los hombros de las bibliotecas universitarias, pero estamos convencidos de que la solución debe pasar por contar con ellas, con su producción y con el trabajo que han realizado para describir los contenidos que poseen. Que esa solución se desarrolle de manera aislada sólo conllevaría la creación de un portal más de los muchos que existen dedicados a estas tareas.

Se propone el uso de un cms para la gestión de toda esa información. Un cms con soporte semántico, que sea capaz de aportar más que un simple repositorio institucional, y que permita a sus usuarios interactuar, crear redes sociales en función de afinidades temáticas y/o profesionales, compartir documentación y resultados de sus investigaciones,



Figura 5. La fase piloto de *Nexus* se está probando en la *Universidad Central Marta Abreu* de Cuba bajo el nombre de *Scientia*

conectar las aportaciones de cada uno con los fondos de la biblioteca y, como colofón de todo eso, aumentar la visibilidad de personas, institución y biblioteca.

El software propuesto, *Nexus*, se perfila como una herramienta para el desarrollo de este tipo de servicios de información usando para ello tecnologías semánticas, sin que ello merme las posibilidades de comunicación y conexión entre académicos de diversas entidades educativas y científicas. Las ventajas competitivas que ofrece *Nexus* sobre el resto de redes sociales académicas que hoy existen reside en proporcionar una mayor interoperabilidad entre los sistemas (e información) ya existentes en la biblioteca universitaria, facilitar el control bibliográfico y la gestión, y acceder a contenidos que no son recuperables mediante catálogos.

6. Bibliografía

- Allemang, Dean; Hendler, James** (2011). *Semantic web for the working ontologist: effective modelling in RDFS and OWL*. 2ª ed. Morgan Kaufmann. ISBN: 978 0123735560
- Alonso-Sierra, Luis-Enrique; Ortiz-Muñoz, Ernesto; Hidalgo-Delgado, Yusniel** (2012). "Los sistemas de gestión de contenidos en el ámbito de la web semántica: una breve revisión". *Serie científica de la Universidad de las Ciencias Informáticas*, v. 5, n. 11, pp. 1-9.
<http://publicaciones.uci.cu/index.php/SC/article/view/945>
- Amante, Maria-João; Extremeño, Ana** (2012). "Bibliotecarios universitarios - profesores. ¿Caminos convergentes?". *Revista española de documentación científica*, v. 35, n. 2, pp. 298-324.
<http://redc.revistas.csic.es/index.php/redc/article/view/738/819>
- Barrueco-Cruz, José-Manuel; García-Testal, Cristina** (2009). "Repositorios institucionales universitarios". *XI Jornadas españolas de documentación*: 20-22 de mayo de 2009, pp. 99-108. ISBN: 84 692 2663 6
<http://www.fesabid.org/zaragoza2009/actas-fesabid-2009/99-107.pdf>
- Breslin, John G.; Decker, Stefan; Harth, Andreas; Bojars, Uldis** (2006). "SIOC: an approach to connect web-based communities". *Intl journal of web based communities*, v. 2, n. 2, pp. 133-142.
http://www.johnbreslin.org/files/publications/20060610_ijwb2006.pdf
- Byron, Angela; Berry, Addison; De-Bondt, Bruno** (2012). *Using Drupal: choosing and configuring modules to build dynamic websites*. 2ª ed. Sebastopol (California): O'Reilly Media. ISBN: 978 1 4493 9052 5
- Chuttur, Mohammad Y.** (2012). *Training and best practice guidelines: implications for metadata creation*. Indiana University.
- Domínguez-Velasco, Sandor; Senso, José A.; Leiva-Mederos, Amed** (2011). *Puertotex*. Software en versión beta. Santa Clara, Cuba: Cdiect, Universidad Central de las Villas.
- Domínguez-Velasco, Sandor** (2010). *Colector: herramienta para exportar términos a SKOS, FOAF y RDF*. Versión alfa. Depto. de Investigación y Desarrollo de Centro de Documentación e Información, Univ. Central de las Villas, Santa Clara, Cuba.
- Garrido, Piedad; Tramullas, Jesús** (2006). "Fundamentos". En: Tramullas, Jesús; Garrido, Piedad (coords.). *Software libre para servicios de información digital*. Madrid: Prentice Hall, pp. 1-22. ISBN: 978 8483222997
- Hebrang-Grgić, Ivana** (2011). "Usage of open access institutional repositories in some of the European peripheral scientific communities". En: Hemera, Annette; Mateljan, Vladimir; Banek Zorica, Mihaela; Stančić, Hrvoje; Seljan, Sanja (ed). *InFuture 2012. The future of information sciences*. Zagreb: Dept of Information Sciences, pp. 33-42. ISBN: 953 175 408 8
http://hmk.ffzg.hr/bibl/InFuture2011/PDF/01_e-Society/1-04_Hebrang_Grgic_Usage_of_Open_Access_Institutional_Repositories.pdf
- Leiva-Mederos, Amed; Domínguez-Velasco, Sandor; Senso, José A.** (2012). "Puertotex: un software de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías". *Transinformação*, v. 24, n. 2, pp. 103-115.
<http://www.citeulike.org/user/jsenso/article/11747389>
- Library of Congress (2011). *Library of Congress subject headings: linked data service*.
<http://id.loc.gov>
- Lillevig, Grace; Stevens, Linda** (2009). "Community conversations through online tools". *Texas library journal*, v. 86, n. 2, pp. 55-57.
- Melero, Remedios; Abadal, Ernest; Abad-García, Francisca; Rodríguez-Gairín, Josep-Manuel** (2009). "The situation of open access institutional repositories in Spain". *Information research*, v. 14, n. 4.
<http://informationr.net/ir/14-4/paper415.html>
- Méndez, Eva; Bravo, Alejandro; López, Leandro** (2007). "Microformatos: web 2.0 para el Dublin Core". *El profesional de la información*, v. 16, n. 2, pp. 107-113.
<http://www.elprofesionaldelainformacion.com/contenidos/2007/marzo/03.pdf>
<http://dx.doi.org/10.3145/epi.2007.mar.02>
- Merlino-Santesteban, Cristian** (2012). "Repositorios institucionales y buscadores web. Una interrelación no tan exitosa". En: *10ª Jornada sobre la biblioteca digital universitaria*, 8-9 nov., Buenos Aires.
<http://nulan.mdp.edu.ar/1664/1/01365.pdf>
- Merlo-Vega, José-Antonio; Angosto-Castro, Antonia; Ferreras-Fernández, Tránsito; Maestro, José-Ángel; Gallo-León, José-Pablo; Ribes-Llopes, Inmaculada; Marcos, Amada** (2011). *Science 2.0: the use of social networking in research*. Madrid: Rebiun.
<http://eprints.rclis.org/16162>
- Mooney, Sean D.; Baenziger, Peter H.** (2007). "Extensible open source content management systems and frameworks: a solution for many needs of a bioinformatics group". *Briefings in bioinformatics*, n. 9, pp. 68-74.
<http://bib.oxfordjournals.org/content/9/1/69.full.pdf+html>

Pastor-Sánchez, Juan-Antonio (2011). "Drupal 7: web semántica al alcance de todos". En: *XII Jornadas españolas de documentación. XVI Jornadas bibliotecarias de Andalucía*. Málaga: Fesabid 11.

<http://digitum.um.es/xmlui/handle/10201/21332>

Pastor-Sánchez, Juan-Antonio (2012). "Los cms como pieza fundamental en el despliegue de la web semántica". *Anuario ThinkEPI*, v. 6, pp. 184-189.

Phipps, Jon; Hillman, Diane (2011). "The open metadata registry: an update". *Bulletin of the American Society for Information Science and Technology*, v. 37, n. 4.

http://www.asis.org/Bulletin/Apr-11/AprMay11_Phipps_Hillmann.pdf

Poornima, Narayana (2012). *Design & development of ins-*

titutional repositories: problems and perspectives. Lambert Academic Publishing, Germany. ISBN: 978 3 8473 7652 1

Sánchez, Adam (2011). *Drupal: "Agrovoc Field" module to index nodes with Agrovoc*.

<http://aims.fao.org/fr/community/agridrupal/pages/drupal-agrovoc-field-module-index-nodes-agrovoc>

Schöpfel, Joachim; Bescond, Isabelle; Prost, Hélène (2012). "Open is not enough: a case study on grey literature in an OAI environment". *The grey journal*, v. 8, n. 2, pp. 112-124.

Zamora, Raquel (2012). "La evolución de Biblioteca a Centro de Recursos de Aprendizaje (CRAI)". *Universidad y sociedad*, v. 4, n. 1, pp. 1-8.

<http://www.ucf.edu.cu/ojsucf/index.php/uys/article/view/139>

Colección de libros de bolsillo

EPI-UOC



Más información:

<http://www.elprofesionalde lainformacion.com/libros.html>