



# Integración de Repositorios Digitales para la Gestión del Conocimiento en el ámbito universitario Colombiano

Luis Eduardo Cano Olivera <sup>1</sup>  
 Pedro Enrique Espitia Zambrano <sup>2</sup>  
 José Nelson Pérez Castillo <sup>3</sup>  
 Luis Joyanes Aguilar <sup>4</sup>

## Resumen

Este artículo presenta los pasos para realizar los cambios de integración y migración de información de los repositorios digitales en diversas plataformas de código abierto (Open Source) con el fin de garantizar la estandarización de las soluciones informáticas propias de las Instituciones de Educación Superior (IES). Además de mejorar el acceso a la aplicación impidiendo que personas no autorizadas a la IES se registren dentro del sistema.

Adición a este problema, existe el que se viene presentando dentro de las Tecnologías de la Información y las Comunicaciones (TICs), con el auge de información que es registrada segundo a segundo a nivel mundial por el fenómeno de la Web 2.0 [9]. Información que en la mayoría de los casos no puede ser reutilizada tanto por centros de investigación y/o sector académico al no existir adecuadas herramientas informáticas que permitan centralizar la gestión del conocimiento en un determinado sitio y de manera especializada según los diferentes Núcleos Básicos del Conocimiento (NBC) que existen dentro del sector educativo [5], sin olvidar además los costos en las licencias de software que puedan existir.

Pese a que los movimientos del software libre [6][7] y del código abierto [8], los que no se deben confundir, hayan suministrado algunas herramientas informáticas para la gestión del conocimiento; en la mayoría de los casos estas herramientas no se pueden integrar entre sí con otros sistemas de información que tenga una IES, debido a la poca documentación que existe y/o a la especificación única dada a un determinado lenguaje de programación y/o motor de base de datos y/o plataforma tecnológica con que fueron desarrollados, entre otros. Obligando a que en una determinada IES tenga que realizar diferentes implementaciones descentralizadas con información redundante en sus sistemas de información, sin que éstas estén sincronizadas con otras soluciones informáticas que tenga la misma institución.

**Palabras clave:** Conocimiento, Repositorio Digital, Sistema de Información, Software Libre, Software Código Abierto, Biblioteca Digital, Objetos de Aprendizaje, Institución de Educación Superior.

## Integration of Digital Repositories for the Knowledge Management in the Colombian University Ambience

### Abstract

This article presents the steps to realize the changes of integration and migration of information of digital repositories in different platforms of Open Source (Open Source) in order to guarantee the standardization of any proper informatic solutions of the Universities of High School (UHS). In addition to improving the access to the application preventing persons not authorized to the UHS from registering inside the system.

Addition to this problem, it exists the one that has appearing inside the Information Technologies and the Communications (ITCs), with the heyday of information that is registered second to second on a global scale by the phenomenon of the Web 2.0 [9]. Information that in most cases cannot be re-used so much by research centers and/or academic sector after there did not exist suitable informatic tools that allow to centralize the knowledge management in a certain place and of way specialized according to the different Basic Nuclei of the Knowledge (BNK) that exist inside the educational sector [5], without forgetting also the costs in the licenses of software that could exist.

Although the movements of the free software [6] [7] and with the open source [8], which must not make a mistake, they have provided some informatic tools for the knowledge management; in most cases this tools cannot integrate between themselves with other information systems who has an UHS, due to few documentation that exists and/or to the only specification started to a certain computer language and/or engine of database and/or technological platform with which they were developed, between others. Forcing that in a certain UHS has to realize different implementations decentralized with redundant information in his information systems, without these being synchronized with other informatic solutions that the same institution has.

**Key words:** Knowledge, Digital Repository, Information System, Free Software, Open Source Software, Digital Library, Objects of Learning, University High School.

<sup>1</sup> Doctorando en Informática de la Universidad Pontificia de Salamanca, España. Miembro del Grupo de Investigación GICOGE.

<sup>2</sup> Consultor en Branding del Tecnológico de Monterrey sede Colombia. Estudiante de Doctorado en Informática de la Universidad Pontificia de Salamanca de España. Miembro del grupo de Investigación en Tráfico Inteligente y Seguridad Ciudadana.

<sup>3</sup> Ph.D. en Informática Universidad de Oviedo. Profesor Universidad Distrital, Facultad de Ingeniería. Investigador principal Grupo de Investigación GICOGE.

<sup>4</sup> Ph.D. en Informática de la Universidad de Oviedo. Ph.D. en Ciencias Políticas y Sociología de la Universidad Pontificia de Salamanca. Miembro del grupo GICOGE. Director del grupo de investigación: Ingeniería del Software y Sociedad de la Información y el Conocimiento, Universidad Pontificia de Salamanca.

## 1. INTRODUCCIÓN

En el desarrollo de los programas académicos y de los centros de investigación que ofrecen las Instituciones de Educación Superior (IES), se generan constantemente muchos datos, resultado de proyectos de investigación, trabajos, monografías y tesis, entre otros. Sin olvidar que de toda esta cantidad de datos, se establece que entre un 60% y 70% se convierten en información y que de este rango, sólo entre un 5% y 15% se convierte en verdadero conocimiento. También se destaca, que de ese porcentaje de datos, en su mayoría desconocido por muchos sectores, conlleva a que se repitan y/o dupliquen una y otra vez en diferentes proyectos con diferentes nombres los mismos contenidos, presentando estancamiento en la Gestión del Conocimiento.

Es en este sentido que surge la pregunta, *¿En dónde puede existir un sistema dentro del sector educativo de consulta de resultados y/o avances de proyectos ya sean fallidos y/o exitosos, para que se tomen como referencia para nuevos prototipos y/o proyectos de investigación?* La respuesta es objetiva, dentro del sector bibliotecario de una IES, el cual, está encargado de almacenar y gestionar toda la producción intelectual que le llegue, con un pequeño inconveniente detectado y es que ésta no queda digitalizada sino en

físico. Además, si le agregamos el factor de desplazamiento geográfico, se puede deducir que alguien interesado en un tópico diferente no puede tener acceso a la información de su interés y tendría que empezar desde cero sus estudios sin poder reutilizar los resultados previos al área de interés de cualquier investigador.

En razón a ello, ha surgido una solución para este tipo de situaciones proveniente de las Tecnologías de la Información y las Comunicaciones (TICs) conocidas como Repositorios, algunos de licencia propietaria (Copyright) y otros de software libre (GPL) o de código abierto (Open Source), los cuales no se deben confundir entre sí, siendo una alternativa para las IES que quieran incursionar en este campo de acción, pero que por sus características técnicas, no son fáciles de integrar con otro tipo de Sistemas de Información (SI). En éste escrito se explica metódicamente, cómo pueden las IES enfrentar estos

inconvenientes, independiente de las características técnicas y/o lógicas que tengan dichas soluciones informáticas para el manejo de los Repositorios.

## 2. DIFERENCIANDO DATOS, INFORMACIÓN, CONOCIMIENTO Y GESTIÓN DEL CONOCIMIENTO

Es importante puntualizar en los conceptos de: *datos, información, conocimiento y gestión del conocimiento* y cómo están relacionados entre sí, en vista a que se presenta confusión entre estos conceptos llevando a una mala interpretación de los mismos y afectando en última instancia los resultados obtenidos en cualquier implementación que se desarrolle.

**2.1. Dato:** Se conoce como una representación simbólica (*que puede ser numérica, alfabética, algorítmica etc.*), un atributo o una característica de una entidad, el cual, no tiene un valor semántico (*es decir, sentido*) en sí mismo y que son generadas por las personas. En informática, se dice además, que el dato se encuentra dispuesto de una manera adecuada para su tratamiento por un determinado computador. [1][3][4][34]. Convirtiéndose en la base de la información.

**2.2. Información:** Se ha definido como un conjunto de datos que son procesados de una manera organizada y con un determinado sentido semántico, constituyendo un mensaje sobre un ente o entidad. [1][3][4][34].

**2.3. Conocimiento:** Surge a partir del análisis y uso productivo de la misma información, definiéndolo en última instancia como el conjunto organizado de datos e información que permiten resolver un determinado problema o tomar una decisión (*es decir, un conocimiento "accionable"*). [1][3][4][34].

**2.4. Gestión del Conocimiento:** Se ha definido como un proceso tanto cultural como tecnológico, que busca captar y transformar el conocimiento individual para sistematizarlo y convertirlo en *información valiosa* de acceso colectivo. [1][3][4] Ver Figura 1.

## 3. IMPORTANCIA DE LA EDUCACIÓN, LAS UNIVERSIDADES Y LOS NÚCLEOS BÁSICOS DEL CONOCIMIENTO

Otros actores a tener en cuenta en este proceso, son aquellos que determinan las directrices de la educación dentro de los programas de formación profesional ofrecidos como son las Instituciones de

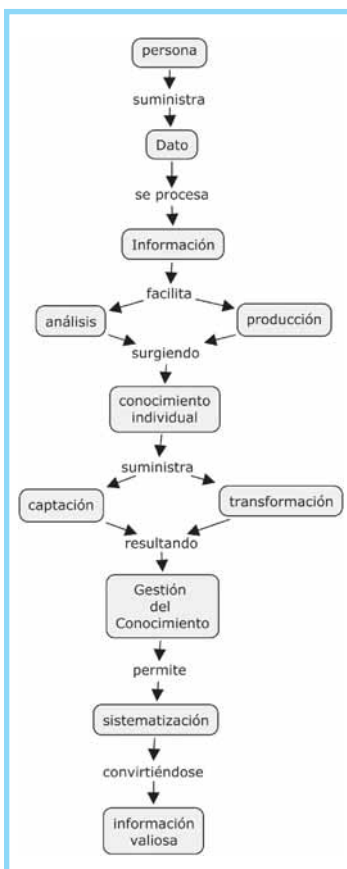


Figura 1. Generación de información valiosa de acceso colectivo. Fuente: adaptada por los autores.

Educación Superior (IES), y que se fundamentan en las diferentes categorías existentes de los Núcleos Básicos del Conocimiento (NBC) [5] suministrados por el Ministerio de Educación Nacional (MEN)[5], para analizar el proceso de formación de la educación nacional en Colombia.

Los “*Núcleos Básicos del Conocimiento*” (NBC), consisten en la afinidad que tengan en los contenidos programáticos y los campos específicos del conocimiento existentes. En total existen 55 NBC, entre los que se destacan: Economía-Administración-Contaduría y afines, la Ingeniería-Urbanismo y afines, las Ciencias Sociales y Humanas, la Agronomía-Veterinaria y afines y la Matemáticas y Ciencias Naturales, según las estadísticas suministradas por el mismo MEN. [5]

Esta categorización de los NBC, ha facilitado el aprendizaje y fortalecimiento de grupos de investigación, semilleros de investigación, comunidades especializadas multidisciplinarias y del sector educativo en general. Pero para que se pueda realizar una congruente Gestión del Conocimiento, se necesita la implementación de tecnologías basadas en repositorios digitales, las cuales están siendo implementadas en Colombia a través de una herramienta de código abierto para repositorios como lo es DSPACE [21] con un 92% de aceptación, pese a que existen otras en el mercado.

Necesidades que en Colombia han sido analizadas y adoptadas por las siguientes IES: Corporación Universitaria Lasallista [36], Corporación Universitaria Minuto de Dios. [37], E-medicina fetal [38], Instituto Interamericano de Cooperación para la Agricultura IICA oficina en Colombia.[39], Universidad Autónoma de Occidente [40], Universidad de Antioquia (UdeA) [41], Universidad de la Salle [42], Universidad de los Andes [43], Universidad del Rosario [44], Universidad ICESI [45], Universidad Nacional de Colombia [46] y Universidad Nacional Abierta y a Distancia de Colombia (UNAD)[47]. Ver figuras 2 y 3.

Y le sigue en segundo lugar, SciELO [52], un modelo para la publicación electrónica cooperativa de revistas científicas en Internet con un 8%, adoptada en la actualidad por la Universidad Nacional de Colombia [46]. (Ver Figura 4).

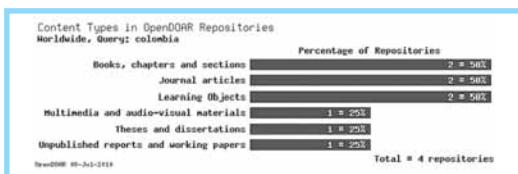


Figura 2. Estadísticas en Colombia sobre Repositorios por el tipo de Contenido. Tomado de OpenDoar [35].

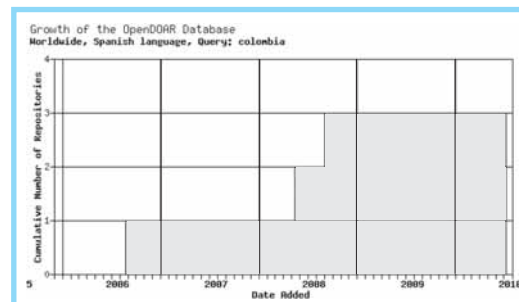


Figura 3. Estadísticas en Colombia sobre el crecimiento de base de datos sobre Repositorios del 2006 al 2009. Tomado de OpenDoar [35]

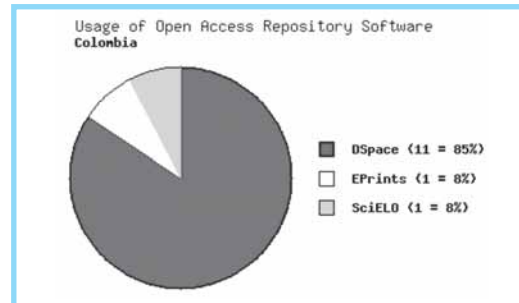


Figura 4. Estadísticas sobre el software para repositorio más utilizado en Colombia. Tomado de OpenDoar [35].

## 4. SISTEMAS DE INFORMACIÓN (SI) Y REPOSITARIOS

### 4.1. Sistema de Información

Cuando se habla de un “*Sistema de Información*”(SI) se nos viene a la mente, el conjunto de elementos que están orientados al tratamiento y la administración de los datos y la información, los cuales, ayudan en la generación de nuevo conocimiento, de acuerdo a su propósito. Debido al crecimiento en datos e información procesados por los sistemas de información para que puedan ser consultados de manera centralizada, ha surgido el concepto de repositorios.

### 4.2. Repositorio

Un “*repositorio, depósito o archivo*” se ha definido como un sitio centralizado donde se almacena y mantiene la información digital, habitualmente en bases de datos o archivos informáticos. En el latín original, la palabra *repositorium* significa armario o alacena. El inglés la asume como *repository* y el castellano como *repositorio*. El término se usa actualmente para definir los archivos abiertos, de ámbito institucional y de acceso público, que almacenan trabajos en soporte digital. [57].

El *Diccionario de la lengua española* de la Real Academia Española de 2001, incluye la palabra *repositorio* como: “lugar donde se guarda una cosa”. La última edición

del *Diccionario Collins universal* de 2005 la traduce como: “depósito”. [58]. En vista de este significado el imaginario colectivo suele hacerse a la idea mental de un depósito como un almacén o armario donde se puede guardar algo. La palabra *repositorio en este contexto* connota a algo archivado, escondido, inmovilizado; recuerda algo antiguo y pasado de moda. [57]. Todo ello inconsciente, que podría agregarse a los doce factores que apunta Alice Keefer (2007) que explican la resistencia de muchos autores a incluir sus producciones intelectuales en los repositorios. [59].

Estas apreciaciones quedan lejos del dinamismo que la palabra le exige al significado moderno, que hace referencia a un espacio institucional, dinámico, donde se puede incluir un documento con la certeza de que quedará a disposición de un abanico amplio de usuarios. Es así que los “depósitos” están preparados para distribuirse habitualmente sirviéndose de una red informática como Internet o en un medio físico de información digital. Y pueden ser de acceso público, o pueden estar protegidos y necesitar de una autenticación previa. Los depósitos más conocidos son los de carácter académico e institucional.

A diferencia de los computadores personales (PCs), los depósitos suelen contar con sistemas de Backup y mantenimiento preventivo y correctivo, lo que hace que la información se pueda recuperar en el caso de que una máquina o PC quede inutilizable. Los depósitos se utilizan de forma intensiva en el sistema operativo GNU-Linux, almacenando, en su mayoría paquetes de software disponibles para su instalación mediante un gestor de paquetes.

### 4.3. Clasificación de Repositorios:

Dentro de los repositorios han surgido los siguientes, de acuerdo al enfoque al que están dirigidos:

#### 4.3.1. Repositorio Digital (RD):

Un sistema en red formado por hardware, software, data y procedimientos que:

- Contiene objetos digitales (OD).
- Contiene metadatos.
- Asegura la identificación persistente del objeto mediante un identificador único persistente.
- Ofrece funciones de gestión, archivo y preservación de los objetos.
- Proporciona un acceso fácil, controlado y estandarizado a los objetos.
- Ofrece los sistemas adecuados de seguridad para los objetos y los metadatos.
- Sostenible en el tiempo.

El contenido del RD permite que se conserve:

- Patrimonio cultural de las organizaciones
- Repositorios académicos
- Documentos de organizaciones gubernamentales
- Literatura gris
- Documentos, folletos, boletines, presentaciones, conferencias y otros tipos de materiales.

#### 4.3.2. Repositorio de Objetos de Aprendizaje (ROA)

Son un catálogo electrónico/digital que facilita las búsquedas en Internet de objetos digitales para el aprendizaje. [13]. Se caracteriza porque tienen:

- Búsqueda
- Recuperación/Descarga
- Almacenamiento
- Publicación
- Colectación

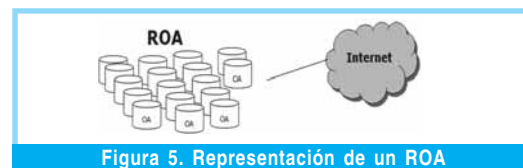


Figura 5. Representación de un ROA

#### 4.3.3. Repositorio Temático (RT)

El RT reúne sus contenidos en función de su área temática, no por su origen institucional (ej. Arxiv) en torno a una disciplina científica o disciplinas relacionadas.

#### 4.3.4. Repositorio Institucional (RI):

Es aquel creado, mantenido y autorizado por una Institución (no exclusivamente una Universidad) o un grupo de Instituciones, que recoge los contenidos digitales generados por la actividad de los miembros de esa Institución. [49]. La organización SPARC [50] define los Repositorios Institucionales como sigue:

- Pertenecen a una institución
- Son de ámbito académico.
- Son acumulativos y perpetuos.
- Son abiertos e interactivos.

Clifford Lynch, refiriéndose al ámbito universitario define el repositorio institucional como “un conjunto de servicios que una Institución ofrece a su comunidad para la gestión, y difusión de los contenidos digitales generados por los miembros de esa comunidad. Es, en su nivel más básico, un compromiso organizativo para el control de esos materiales digitales, incluyendo su preservación, su organización, acceso y distribución” [48]. El tipo de contenido que puede manejar el RI es:

- Documentos “textuales”: libros, tesis, pre-prints, artículos de libros, ponencias, documentos de trabajo, informes técnicos, revistas, entre otros.
- Objetos de aprendizaje (OA)
- Software
- Imágenes, estáticas y en movimiento

- Aplicaciones Multimedia
- Simulaciones
- Páginas web
- Blogs
- Posters, presentaciones , diapositivas, entre otros.

#### 4.3.5. Repositorio Académico (RA)

Reúne únicamente los documentos resultantes de la actividad académica de los miembros de una IES, en concreto los proyectos fin de carrera, grado, monografías, máster o doctorado.

#### 4.4. Software para Repositorios Digitales

De acuerdo al tipo de licenciamiento, se han desarrollado aplicaciones para Repositorios, destacándose las siguientes:

##### 1. Con licenciamiento comercial (Copyright):

- Existen:
- CONTENTdm [14]
  - Digital Commons [15]
  - DigiTool [16]
  - EQUILLA [17]
  - intraLibrary [18]
  - Open Repository [19]
  - VITAL [20]

Aunque tengan buenas características, entre las desventajas presentan están el costo por licenciamiento. Además, de que el código fuente es cerrado y con restricciones, por lo tanto no se pueden integrar con otros sistemas de información. Dependen de un solo proveedor en el aplicativo, y si se quiere integrar con otras soluciones informáticas, las IES deben pagar por dicha implementación.

##### 2. Con licenciamiento Freeware:

- Existe:
- Zenty de Microsoft[29]
- Aunque este tipo de software sea gratis, es decir, no se pague por el licenciamiento y su uso sea ilimitado, la desventaja es que el código es cerrado siendo difícil de integrar con otras soluciones informáticas. Además que dependen de una sola plataforma tecnológica como lo puede ser el Sistema Operativo Windows sin ser multiplataforma.

##### 3. Con licenciamiento Código Abierto (Open Source):

- Existen:
- DSpace [21]
  - E-Prints [22]
  - Fedora Commons [23]
  - Greenstone[24]
  - CDSware [25]
  - Connexions [26]
  - PlanetDR [27]
  - DOOR [28]

Se caracterizan porque el software está orientado a los beneficios prácticos de compartir el código, dando

a los usuarios la libertad de mejorarlos [8], además que no contempla restricciones sobre su uso o modificación y que no se deben confundir con el software de licenciamiento libre. [6][7]. Este último grupo de licenciamiento, es el que corresponde al enfoque del presente escrito.

## 5. COMPARACIÓN ENTRE LOS PROGRAMAS PARA ROAS DE CÓDIGO ABIERTO

El sitio Repositories Support Project (RSP) [10] realizó una encuesta sobre los once (11) repositorios más utilizados a nivel mundial en Marzo del 2009 [30]. Analizando las características fundamentales de cada software incluyendo aspectos como: los tipos de ítems que soporta, interface de usuario, validación de usuarios, plataformas de software, interoperabilidad, funciones de administrador, ayuda, documentación y servicios. [30]

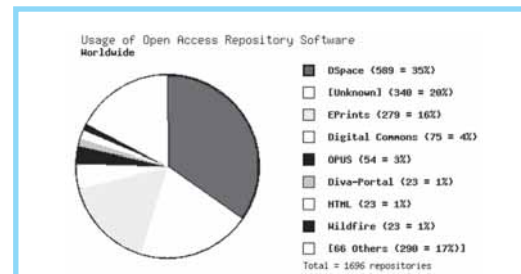


Figura 6 Comparación de diferentes herramientas para repositorios. Fuente RSP [10]

De ese estudio hemos tomado como referencia únicamente los programas de Open Source: DSpace, Eprints, Fedora Commons y PlanetDR, constituyéndolos en el objeto de estudio del presente documento. De estas herramientas nos concentraremos únicamente en aquellas características que son vitales para la integración de los repositorios con los sistemas de información ó que tiene algún tipo de incidencia al respecto:

### 5.1. Validación del Usuario:

- a) **El registro es necesario** en estas aplicaciones, estando implementado de manera estándar a excepción de Fedora Commons que no lo contempla.

En DSpace, el usuario es el correo electrónico, y no permite a múltiples usuarios usar la misma dirección de correo electrónico. Siendo una desventaja porque un usuario se puede registrar varias veces con cuentas de correo diferentes. El sistema provee una característica de recordatorio de contraseña.

En cambio, con las aplicaciones de los Eprints, Fedora Commons y PlanetDR, el sistema requiere un usuario y contraseña para registrarse, además una

dirección de correo electrónico para utilizarla en caso de que su contraseña sea olvidada. Varios usuarios pueden usar la misma dirección de correo electrónico, siendo un inconveniente porque no se puede garantizar el acceso de un usuario único con una sola cuenta de correo única, con el agravante de suplantación de identidad.

b) **Contemplación de la utilización LDAP** (*Lightweight Directory Access Protocol, Protocolo Ligero de Acceso a Directorios*), que consiste en almacenar la información de autenticación (*usuario y contraseña*) para autenticarse aunque sea posible almacenar otro tipo de información (*datos de contacto del usuario, ubicación de diversos recursos de la red, permisos, certificados, etc*). La aplicación que lo utiliza de manera estándar es Eprints siendo opcional en las otras.

c) **Utilización de la autenticación Atbeans (Ate-nas)**: Contempla un mecanismo de autenticación para conceder el acceso a la información en línea de los servicios electrónicos (*recursos digitales*) adquiridos por una IES.

d) **Utilización de la autenticación Shibboleth**: Es empleada para resolver las necesidades de cada organización para el intercambio de información sobre sus usuarios de manera segura. A través de la arquitectura Shibboleth, cada aplicación solicita la autenticación de los usuarios y los autoriza para usar funciones específicas o hacer búsquedas que utilicen recursos específicos para quienes estén autorizados. Esta herramienta viene incluida para ser habilitada de manera opcional en las aplicaciones de estudio.

e) **No contienen módulos para facilitar la administración de usuarios y contraseñas de manera automática o para importarlos de otros Sistemas de Información**: Pese a que estos sistemas informáticos de repositorios digitales tienen buenas estructuras en seguridad manejando algoritmos de encriptación de contraseñas como MD5, no cuentan con un mecanismo que permita migrar y/o importar los usuarios y contraseñas que maneja un aplicación propia de la IES, y registrarlos dentro de las mismas aplicaciones de repositorios, obligando a cada persona que se tenga que registrar en dicho sistema, las veces que quiera o sea necesario de acuerdo al número de comunidades a las que desee acceder.

## 5.2. Plataforma Tecnológica

a) **Sistema Operativo**: Estas soluciones funcionan de manera mono-plataforma o multi-plataforma. Siendo recomendable el uso de aquellas que sean multi-plataforma, con el fin de ahorrar costos a las IES

que estén trabajando en una plataforma con un sistema operativo específico y que no tengan inconvenientes de migración. La mayoría de sistemas informáticos para repositorios se orientan al Sistema Operativo Windows. Pero se recomienda que por lo menos el software sea compatible además con las plataformas Mac, GNU-Linux, y Unix y sus derivados. Ver Figura 7.

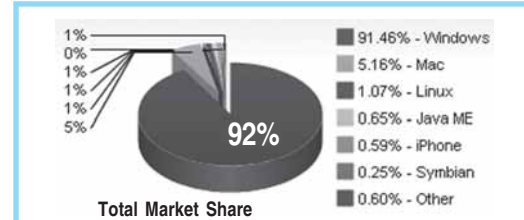


Figura 7. Comparación de Sistemas Operativos más utilizados. Fuente Marketshare [55]

b) **Bases de Datos (BD)**: Estas herramientas están adecuadas para un tipo de base de datos específico. En el momento de la instalación tienen una serie de scripts de creación de base de datos de acuerdo al motor de BD que soporte. Por ejemplo, MySQL, no es soportada por DSPACE [21], a diferencia de EPrints[22], FedoraCommons[23] y PlanetDR[27] que si la contemplan. En el caso de Oracle, es solamente contemplado por DSPACE[21] y EPrints[22]. O el caso de PostgreSQL que la contempla solo DSpace[21] y FedoraCommons[23]. Recordando por último, que estas no son las únicas bases de datos que existen. Ver Figura 8.

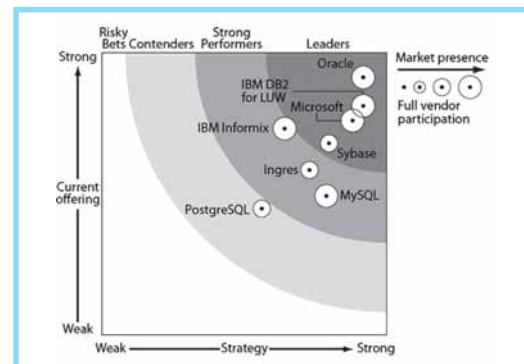


Figura 8. Comparación de Gestores de Bases de Datos. Fuente Forrester [56]

En este aspecto, es en donde se presenta la limitante para integrarlo con los gestores de base de datos utilizados por las IES, de acuerdo a los procesos académicos y administrativos que manejen.

c) **Lenguajes de Programación**: Cada sistema de información se encuentra desarrollado en un lenguaje de programación específico, de acuerdo a las necesidades suministradas por la entidad. Las IES cuentan con soluciones informáticas propias desarrolladas por algún lenguaje llámese: php, jsp, asp, vbs u otro, y deben someterse al tipo de lenguaje en el que haya sido desarrollada la herramienta de software código libre. Ver tabla I.

**Tabla I. Estadísticas de los lenguajes de programación.**  
Fuente: Tiobe [11]

Position Jun 2010	Position Jun 2009	Delta in Position	Programming Language	Ratings Jun 2010	Delta Jun 2009	Status
1	1	==	Java	18.033%	-2.11%	A
2	2	==	C	17.800%	+1.03%	A
3	3	==	C++	10.757%	+0.16%	A
4	4	==	PHP	8.934%	-0.74%	A
5	5	==	(Visual) Basic	5.868%	-2.07%	A
6	7	↑	C#	5.196%	+0.66%	A
7	6	↓	Python	4.266%	-0.49%	A
8	9	↑	Perl	3.200%	-0.71%	A
9	45	↑↑↑↑↑↑↑↑	Objective-C	2.469%	+2.35%	A
10	11	↑	Delphi	2.394%	+0.21%	A
11	8	↓↓	JavaScript	2.191%	-1.83%	A
12	10	↓↓	Ruby	2.070%	-0.56%	A
13	12	↓	PL/SQL	0.787%	-0.09%	A
14	14	==	SAS	0.703%	-0.06%	A
15	15	==	Pascal	0.702%	-0.06%	A-
16	18	↑↑	Lisp/Scheme/Clojure	0.654%	+0.05%	B
17	19	↑↑	Lua	0.592%	+0.04%	B
18	20	↑↑	MATLAB	0.589%	+0.06%	B
19	16	↓↓	ABAP	0.577%	-0.15%	B
20	27	↑↑↑↑↑↑	PowerShell	0.529%	+0.23%	B

Por ejemplo, DSpace[21], PlanetDR[27] y FedoraCommons[23], basados en tecnología Java, están desarrollados en JSP. Eprints[22], está desarrollado en Perl, utilizando tecnología Ajax y Javascript.

El inconveniente se presenta, cuando el instalador y/o administrador y/o programador de la solución informática no conoce o no domina el lenguaje de programación con el cual fue desarrollada dicha aplicación, empleando una frase de cajón “*eso no se puede hacer o desarrollar o implementar.*”, ocasionando que la IES se quede sin adecuarlo a sus propios requerimientos.

**d) Servidores Web:** Estos sistemas de información están orientados a la Web y dependen del lenguaje de programación con el que fueron desarrollados, destacándose Internet Information Server (IIS), Apache, Tomcat y Jboss, entre otros.

El inconveniente se presenta únicamente para aquellas herramientas que funcionan con tecnología JSP y Java, y que dependen de la compilación del servidor Apache Ant, como es el caso de DSpace[21], cuando se realiza algún cambio en el código fuente de la aplicación, presentando algunos problemas de tipo técnico al respecto.

**e) Formato de los Metadatos:** Todos contemplan los formatos Dublin Core y Qualified DC. A excepción de FedoraCommons[23], que incluye además de los anteriores, los formatos METS, MARC y puede esperar cualquier otro formato. El inconveniente se presenta en la integración con el uso de metadatos nuevos no soportados.

**e) Documentación Técnica:** Todas las herramientas informáticas cuentan con una documentación técnica básica y/o estándar. Pero ninguna profundiza con detenimiento sobre la solución y se concentran en aspectos como instalación, configuración y administración de la solución. Además que no suministran el modelo Entidad-Relación.

**f) Soporte Técnico:** todas las soluciones informáticas analizadas, tienen foro, wiki, o lista de correos, y aclaran únicamente procesos de tipo genérico, más no puntual cuando se requiere integrar alguna solución o mejora al respecto. Dejando esta actividad al programador para que estudie el código y realice las modificaciones necesarias, y enseguida las publique, basándose en la filosofía del código abierto (Open Source) [8].

## 6. INTEGRACIÓN DE LAS APLICACIONES PARA RDS CON LOS SIS INSTITUCIONALES DE LAS IES

Cuando se encuentran los inconvenientes antes mencionados, las IES pierden el control de la solución informática y no pueden realizar procesos de tipo administrativo como autorizar solo a aquellos usuarios (*administrativos, docentes, investigadores, estudiantes, egresados*) que formen parte de dicha entidad, porque el sistema permite el registro de cualquier persona ajena a la IES, o que se encuentre desvinculada de la misma. Además, que permite la duplicidad en la información sin tener un control en los datos básicos del usuario.

La metodología que se explica a continuación, son el resultado del estudio, dedicación e investigación en los conceptos de repositorios, licenciamiento, dominio sobre las herramientas especializadas en repositorio existentes en código abierto, programación e implementación que se realizó en una IES que necesitó dicha integración; las cuales no se encuentran soportadas en ninguna de las páginas oficiales de repositorios a nivel general, ni en los mismos soportes y/o documentos técnicos suministrados por los autores de esas herramientas, con el fin de que cualquier IES los tome como referencia para implementarlos dentro de su organización.

### 6.1. Pasos para integrar una aplicación de Repositorio con un SI propio de una IES

Como resultado de esta investigación se presentan a continuación los pasos para estandarizar repositorios multiplataforma y así poder integrar una aplicación que utilice una IES en las áreas de Registro y Control, o Procesos Académicos de los Programas. Ver Figura 9.

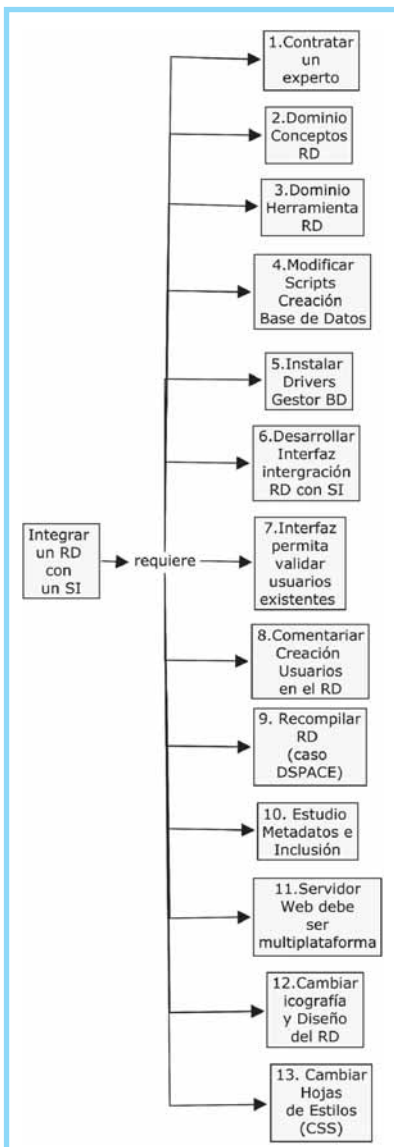


Figura 9. Metodología para integrar un Repositorio Digital con un Sistema de Información. Fuente: los autores

### 1. Contratar un Experto:

Es necesario contratar los servicios de un desarrollador y/o programador informático experto, que tenga conocimiento y/o dominio tanto de las aplicaciones desarrolladas por la IES como de la solución de repositorio a implementar.

### 2. Dominio conceptos de RD:

El desarrollador y/o programador debe tener claro los conceptos de Repositorios y Metadatos.

### 3. Dominio herramienta de RD:

Se debe instalar la aplicación del repositorio digital y conocer todo el funcionamiento de la misma. Dicha documentación se encuentra en la página del autor de la herramienta. Si tiene dudas puede recurrir al foro del autor.

### 4. Modificar scripts Creación Base de Datos:

Todas las aplicaciones estudiadas cuentan con unos scripts para la creación de las tablas que necesita el sistema de repositorio digital.

Si se quiere montar en un motor de base de datos específico que la IES utilice y que no esté soportado por

dicha aplicación de repositorio, se debe buscar un archivo con extensión .sql en donde figure la sentencia Create Database y Create Table, y modificarlo de acuerdo al motor deseado. Estos scripts manejan la creación de campos consecutivos o

autonuméricos que se definen de manera diferente según el estándar ANSI de SQL que contempla cada gestor de base de datos. Para adecuarlo a un motor de base de datos como Microsoft SQL Server, se debe utilizar el lenguaje Transact-SQL, y modificarlo en su totalidad, debido a que su estructura es completamente diferente.

### 5. Instalar Drivers del Gestor de BD:

Se deben buscar los drivers del gestor de base de datos a adecuar en el aplicativo del repositorio digital, los

cuales siempre se encuentran en la página oficial del respectivo motor de base de datos.

### 6. Desarrollar Interfaz de integración RD con SI:

Para poder automatizar la lista de usuarios con sus respectivas contraseñas, así como las cuentas de correo electrónico que tenga la IES, en cualquiera de sus respectivos sistemas de información, es necesario realizar un programa (o interfaz) en el lenguaje de programación original en el que este desarrollada la aplicación de la IES que contenga la información que se va a migrar, sin olvidar el uso de la instrucción MD5 para encriptar las claves.

Por ejemplo DSpace, está desarrollado en Java, y la aplicación de la IES está en PHP. El desarrollador puede utilizar las instrucciones pg\_connect y sus derivados, para conectarse desde la aplicación de la IES a la tabla e-person de la base de datos PostgreSQL de DSpace y alimentar la información en la tabla de usuarios y contraseñas. Ver figura 10.

### 7. Interfaz permita validar usuarios existentes:

La interfaz a desarrollar debe validar que no se duplique la cuenta de correo y que sólo se utilice una vez. Ver Figura 11.



Figura 11. Ejemplo DSpace[21]. Un usuario se puede registrar con diferentes correos.

### 8. Comentarar Creación de Usuarios en el RD:

En vista a que el código fuente de la aplicación del repositorio se puede modificar, se debe inhabilitar las opciones de registro nuevo de usuarios, colocándole comentarios. Ver Figura 12.

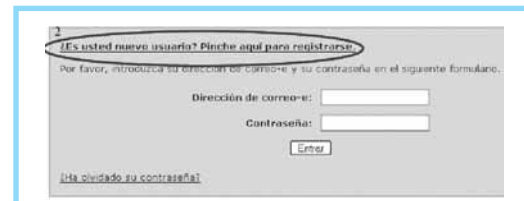


Figura 12. Ejemplo DSpace[21], registro usuarios nuevos. Esta opción no se debe permitir.

### 9. Recompilar RD:

En el caso de DSpace, se debe compilar los cambios realizados con Apache Ant y Apache Maven. Con las otras soluciones informáticas no es necesario este paso. Ver Figura 13.

### 10. Estudio Metadatos e Inclusión:

Para poder incorporar nuevos estándares de metadatos dentro de la aplicación del repositorio no reflejados, se debe estudiar dichos formatos y adicionar las

EPERSON	
EPERSON_ID	
EMAIL	
PASSWORD	
FIRSTNAME	
LASTNAME	
CAN_LOG_IN	
REQUIRE_CERTIFICATE	
SELF_REGISTERED	
LAST_ACTIVE	
SUB_FREQUENCY	
PHONE	
NETID	

Figura 10. Tabla de usuarios de DSpace[21]



respectivas líneas de código en los archivos del código fuente de la aplicación. Ver Figura No 14.

```

Build DSpace and Initialize Database
As the dspace UNIX user, initialize the DSpace database and install DSpace to [dspace]:
od [dspace-source]/dspace/target/dspace-[version]-build.dir
ant fresh_install

[Note] To see a complete list of build targets, run:
ant help

```

Figura 13. Ejemplo DSpace[21], utilización apache ant.

The following configurations allow the administrator extract from the DSpace database a set of records for editing by a metadata export.

Property:	dspace.metadata.valueseparator
Example Value:	dspace.metadata.valueseparator =
Informational:	The delimiter used to separate values within a single field. For example, this will place the double pipe between multiple notes (Seam). This applies to any metadata field that appears more than once in a record. The user can change this to another:
Property:	dspace.metadata.fieldseparator
Example Value:	dspace.metadata.fieldseparator = .

Figura 14. Ejemplo DSpace, utilización metadatos.

**11. Servidor Web debe ser multiplataforma:** Estas aplicaciones están desarrolladas para ejecutarse en un servidor Web multiplataforma. A excepción del Internet Information Server de Microsoft, en donde lo único que se puede configurar son las extensiones de PHP, pero para JSP no se podrían integrar porque requiere de la máquina virtual de Java.

**12. Cambiar iconografía y diseño del RD:** En cuanto al diseño y la iconografía con que vienen estas herramientas informáticas, éstas se pueden cambiar y personalizar según los diseños y los colores corporativos que utilice la IES, formato que viene en jpg y png, y se encuentran ubicadas en una carpeta predeterminada de imágenes.

**13. Cambiar Hojas de Estilo (CSS):** De igual forma, se pueden cambiar las hojas de estilo en cascada (Cascade Style Sheet) que vienen con extensión .css, cambiando los formatos de los fondos, los tipos de letras, colores, enlaces y demás cambios que considere indispensable la IES aplicar de acuerdo a su imagen corporativa.

Realizando los pasos mencionados se puede integrar cualquier solución propia que tenga la IES con las aplicaciones de repositorios digitales.

## CONCLUSIONES

Los pasos para realizar los cambios de integración y migración de información, garantizan que cualquier solución informática propia de la IES se integre con las aplicaciones para repositorios digitales. Además que se puede mejorar el acceso a la aplicación impidiendo que personas no autorizadas a la IES se registren dentro del sistema.

Para realizar cualquier cambio y/o ajuste, se requieren conocimientos avanzados de programación y de conceptualización en repositorios y metadatos.

Aunque DSpace sea la aplicación de más alta difusión en repositorios digitales incluso hasta por la misma comunidad, las otras soluciones de código abierto también son funcionales y se pueden adecuar, mejorar y personalizar de acuerdo a las necesidades de una IES.

Las herramientas repositorios digitales de código abierto no tienen nada que envidiarle a las de licencia propietaria, ofreciendo mejores componentes para la implementación de repositorios digitales de acuerdo a las necesidades que tenga la IES. Además que contribuyen de manera significativa en la Gestión del Conocimiento.

En cuanto a la producción intelectual, existen diferencias en cuanto a dejarlas en copyright (propias para la IES) o en Creative Commons (conocimiento compartido), análisis que sólo le compete a la IES.

Las IES que no adopten este modelo, van a quedar obsoletas dentro de sus procesos administrativos y académicos frente a la comunidad académica; además, que el MEN está exigiendo producción intelectual a las IES y ésta es una de las formas de realizarlo por medio del uso de los repositorios digitales.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] Gestión del Conocimiento <http://www.gestiondelconocimiento.com>
- [2] PRESSMAN, Roger S. INGENIERIA DE SOFTWARE. Un enfoque practico. Sexta Edicion.2005. Editorial Mc Graw Hill
- [3] MICROSOFT. Encarta 2009.
- [4] WIKIPEDIA. <http://www.wikipedia.org>
- [5] Ministerio de Educación Nacional – [www.mineduacion.gov.co](http://www.mineduacion.gov.co)
- [6] STALLMAN, Richard. GNU Operating System <http://www.gnu.org/philosophy/free-sw.es.html>
- [7] STALLMAN, Richard. Free Software Foundation (FSF) <http://www.fsf.org>
- [8] RAYMOND, Eric, Open Source <http://www.opensource.org>
- [9] World Wide Web Consortium (W3C) <http://w3c.org>
- [10] Repositories Support Project <http://www.rsp.ac.uk> . Estadísticas de Repositorios <http://www.rsp.ac.uk/software/surveyresults>
- [11] Tiobe. Estadísticas sobre lenguajes de programación <http://www.tiobe.com>
- [12] Colombia Aprende. <http://www.colombiaprende.edu.co>
- [13] López, C. (2005) Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo). Disponible en: <http://www.biblioweb.dgsca.unam.mx/libros/repositorios/>
- [14] CONTENTdm <http://www.oclc.org/contentdm>
- [15] DigitalCommons <http://www.bepress.com/>
- [16] DigiTool <http://www.exlibrisgroup.com/category/DigiToolOverview>
- [17] EQUELLA <http://www.equella.com>
- [18] IntraLibrary <http://www.intralibrary.com/>
- [19] Open Repository <http://www.openrepository.com/>
- [20] VITAL <http://www.vtls.com/products/vital>
- [21] DSpace, <http://www.dspace.org/>
- [22] Eprints Repository. <http://www.eprints.org/>
- [23] Fedora Commons- <http://www.fedora-commons.org/>
- [24] Greenstone. <http://www.greenstone.org/>

- [25] CERN Document Server Software (CDSware) <http://cdsware.cern.ch/>
- [26] Connexions <http://cnx.org/>
- [27] PlanetDR <http://planet.urv.es/planetdr/>
- [28] DOOR (Digital Open Object Repository) <http://door.sourceforge.net/>
- [29] Zentity <http://research.microsoft.com/en-us/projects/zentity/>
- [30] RSP. Estudio Comparación entre Repositorios. Marzo 2009. <http://www.rsp.ac.uk/software/surveyresults>
- [31] OMPI (ORGANIZACIÓN MUNDIAL DE LA PROPIEDAD INTELECTUAL)-Sistemas automatizados de gestión de derechos y limitaciones y excepciones al derecho de autor. Ginebra. Abril-2006
- [32] Castro Martín, Pablo de. *Estadísticas de uso de repositorios institucionales: ¿un indicador fiable de comparación?* Diciembre 2008, Universidad Complutense de Madrid, Madrid. <http://eprints.ucm.es/8499/>
- [33] DRIVER (Digital Repository Infrastructure Vision for European Research). Directrices DRIVER 2.0. para proveedores de contenido - exposición de recursos textuales con el protocolo OAI-PMH . Noviembre 2008. [http://www.driver-support.eu/documents/DRIVER\\_2\\_1\\_Guidelines\\_Spanish.pdf](http://www.driver-support.eu/documents/DRIVER_2_1_Guidelines_Spanish.pdf)
- [34] REAL ACADEMICA ESPAÑOLA. Diccionario de la Lengua Española. <http://buscon.rae.es/drae/>
- [35] Opendoar. Directorio de Repositorios de Acceso Abierto <http://www.opendoar.org/>
- [36] Corporación Universitaria Lasallista. <http://www.lasallista.edu.co/>. 1. Biblioteca Digital Lasallista (BIDILA). <http://www.lasallistavirtual.edu.co:81/dspace/>
- [37] Corporación Universitaria Minuto de Dios. <http://www.uniminuto.edu/>. 1. Colecciones Digitales Uniminuto. <http://dspace.uniminuto.edu:8000/dspace/>
- [38] e-medicina fetal - <http://e-medicinafetal.org/>. 1. Digital Repository in Fetal Medicine (DRFM). <http://digitalrepository.e-medicinafetal.org/>
- [39] Instituto Interamericano de Cooperación para la Agricultura IICA oficina en Colombia. <http://www.iica.int/colombia>. 1. Centro Virtual de Información CVI IICA oficina en Colombia. <http://zeus.iica.ac.cr:8090/dspace/>
- [40] Universidad Autónoma de Occidente - <http://www.uao.edu.co/>. 1. Biblioteca Digital - Autonomía Virtual. <http://bohr.cuao.edu.co/dspace/>
- [41] Universidad de Antioquia (UdeA) - <http://www.udea.edu.co/>. 1. Biblioteca Digital del Sistema de Bibliotecas de la Universidad de Antioquia. <http://tesis.udea.edu.co/dspace/>
- [42] Universidad de la Salle - <http://unisalle.lasalle.edu.co/>. 1. TEGRA. <http://tegra.lasalle.edu.co/dspace/>
- [43] Universidad de los Andes - <http://www.uniandes.edu.co/1.DSpace> de los Andes. <http://dspace.uniandes.edu.co:5050/dspace/>
- [44] Universidad del Rosario - <http://www.urosario.edu.co/>. 1. edocUR. <http://repository.urosario.edu.co/>
- [45] Universidad Icesi - <http://www.icesi.edu.co/>. 1. Biblioteca Digital - Universidad Icesi. <http://bibliotecadigital.icesi.edu.co/>. 2. Repositorio Académico - Universidad Icesi. [http://repositorioacademico.icesi.edu.co/repositorio\\_academico/](http://repositorioacademico.icesi.edu.co/repositorio_academico/)
- [46] Universidad Nacional de Colombia - <http://www.unal.edu.co/>. 1. Scientific Electronic Library Online - Colombia (SciELO - Colombia). <http://www.scielo.org.co/>. 2. Universidad Nacional De Colombia - Repositorio Institucional UN. <http://www.digital.unal.edu.co/>
- [47] Universidad Nacional Abierta y a Distancia (UNAD). <http://www.unad.edu.co>. Tiene en proceso de implementación varios repositorios en DSPACE. [http://www.unad.edu.co/index.php?option=com\\_content&view=category&id=178&Itemid=938](http://www.unad.edu.co/index.php?option=com_content&view=category&id=178&Itemid=938)
- [48] LYNCH, Clifford A. Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age portal: Libraries and the Academy - Volume 3, Number 2, April 2003, pp. 327-336
- [49] BARTON, Mary. WATERS, Margaret. Cómo crear un Repositorio Institucional. MIT Libraries. 2004-2005 <http://www.recolecta.net/buscador/documentos/mit.pdf>
- [50] SPARC (The Scholarly Publishing & Academic Resources Coalition) <http://www.arl.org/sparc/>
- [51] ROAR. Registry of Open Access Repositories - <http://roar.eprints.org/>
- [52] SciELO- Scientific Electronic Library Online). <http://www.scielo.org>
- [53] López, C. (2005) Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo) <http://www.biblioweb.dgsca.unam.mx/libros/repositorios/>
- [54] OAI. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [55] Marketshare. Comparación en Sistemas Operativos. <http://marketshare.hitslink.com/>
- [56] Forrester. [www.forrester.com](http://www.forrester.com). Comparación de Gestores de Bases de Datos. [http://www.microsoft.com/presspass/italanyst/docs/06-30\\_09\\_Enterprise\\_Database\\_Management\\_Systems.PDF](http://www.microsoft.com/presspass/italanyst/docs/06-30_09_Enterprise_Database_Management_Systems.PDF)
- [57] FRANGANILLO, JORGE. Depósitos: un factor psicológico de desconfianza. Textos universitarios de biblioteconomía i documentació. número 20, juny de 2008. Facultat de Biblioteconomía i Documentació. Universitat de Barcelona, España. <http://www.ub.es/bid/pdf/20frang2.pdf>
- [58] Diccionario Collins universal: español-inglés, inglés-español (2005). Barcelona: Grijalbo.
- [59] Keefer, Alice (2007). "Los repositorios digitales universitarios y los autores". *Anales de documentación*, núm. 10, p. 205-214. <http://www.um.es/fccd/anales/ad10/ad1011.pdf>.

### Luís Eduardo Cano Olivera

Ingeniero de Sistemas de la Universidad Autónoma de Colombia. Especialista en Diseño y Construcción de Soluciones Telemáticas de la Universidad Autónoma de Colombia. Candidato a Magister en Ciencias de la Información y las Comunicaciones con énfasis en Sistemas de Información, de la Universidad Distrital Francisco José de Caldas. Doctorando en Informática de la Universidad Pontificia de Salamanca de España. Docente Universitario. Miembro del Grupo de Investigación GICOGÉ de la Universidad Distrital Francisco José de Caldas y de la línea de trabajo en Geosensores del mismo grupo de investigación. <http://gicoge.udistrital.edu.co/geosensores>. [luiseduardocano@gmail.com](mailto:luiseduardocano@gmail.com)

### Pedro Enrique Espitia Zambrano

Diseñador Gráfico, Universidad Jorge Tadeo Lozano. Especialista en Comunicación de Cultura Científica del CAEU. Editor Senior de Publicaciones Científicas Gaia. Presidente de la Fundación FASARTE, y director del Museo de Arte Vial. Asesor de Greenmax e-technology. Consultor en Branding del Tecnológico de Monterrey sede Colombia. Doctorando en Informática de la Universidad Pontificia de Salamanca de España. Miembro del grupo de Investigación en Tráfico Inteligente y Seguridad Ciudadana de la Universidad Distrital Francisco José de Caldas. [pedroespitia.upsam@gmail.com](mailto:pedroespitia.upsam@gmail.com)

### José Nelson Pérez Castillo

PhD. en Informática Universidad de Oviedo. Profesor Facultad de Ingeniería de la Universidad Distrital Francisco José de Caldas. Magister en Teleinformática Universidad Distrital Francisco José de Caldas. Especialista en Sistemas de Información Geográfica y Teledetección Espacial Universidad de Henares. Investigador Principal Grupo GICOGÉ. [nelsonp@udistrital.edu.co](mailto:nelsonp@udistrital.edu.co)

### Luis Joyanes Aguilar

PhD. en Informática de la Universidad de Oviedo. Doctor en Ciencias Políticas y Sociología de la Universidad Pontificia de Salamanca campus de Madrid. Licenciado en Ciencias Físicas de la Universidad Complutense de Madrid. Licenciado en Enseñanza Superior Militar de la Academia Militar de Zaragoza. Profesor Titular Agregado de Cátedra de Lenguajes y Sistemas Informáticos de la Facultad de Informática de la Universidad Pontificia de Salamanca campus Madrid, donde también se desempeñó como: Decano de la Facultad de Informática, Director de Doctorado, Director de Masters y Director del Departamento de Lenguajes, Sistemas Informáticos e Ingeniería de Software. Miembro del grupo GICOGÉ. Autor de más de 40 libros de Informática y numerosos artículos en revistas y congresos internacionales.