

Propuesta de un Repositorio Administrativo para
Universidad Nacional Experimental del Táchira

**Repositorio de Documentos Administrativos para la Universidad Nacional
Experimental del Táchira (REDAUNET)**

Jose Texier

Profesor de Ingeniería en Informática
Universidad Nacional Experimental del Táchira

jtexier@unet.edu.ve

<http://orcid.org/0000-0003-0176-6625>

Febrero del 2013
San Cristóbal - Estado Táchira - Venezuela

Repositorio de Documentos Administrativos para la Universidad Nacional Experimental del Táchira

Resumen

Los repositorios de documentos digitales administrativos permiten a una organización o institución gestionar la información administrativa de manera electrónica en un formato determinado. Por ello, esta propuesta se centró en establecer las bases para implementar el Repositorio de Documentos Administrativos en la Universidad Nacional Experimental del Táchira (REDAUNET) bajo los principios de la filosofía del Software Libre y del Acceso Abierto. Para fundamentar esta propuesta se realizó un diagnóstico en la UNET mediante un cuestionario cuyos resultados muestran, grosso modo, cómo y qué se hace con los respaldos de los documentos digitales en las instancias de la universidad. Luego, se diseñó el repositorio para la gestión y preservación de los documentos administrativos, a partir del software de repositorios DSpace y adecuado con las leyes del estado venezolano y la normativa de la UNET. REDAUNET cuenta con seis tipos de documentos administrativos: actas, constancias, memorandos, oficios, resoluciones y jurídicos, los cuales contienen distintos metadatos que permiten identificar y localizar los documentos correctamente. La documentación para la instalación, mantenimiento y desarrollo de una plataforma de software para el REDAUNET se generó para servir de apoyo a la propuesta presentada y lograr un producto independiente y de fácil uso.

Palabras Clave: repositorios institucionales, documentos administrativos, UNET, DSpace.

Índice de Contenidos

<u>Resumen</u>	1
<u>Índice de Figuras</u>	4
<u>Índice de Tablas</u>	5
<u>Introducción</u>	6
<u>Capítulo 1</u>	
<u>Consideraciones Iniciales</u>	10
<u>1.1. Situación Actual</u>	10
<u>1.2. Motivación</u>	11
<u>1.3. Objetivos</u>	11
<u>1.3.1. General</u>	11
<u>1.3.2. Específicos</u>	11
<u>1.4. Alcance y Limitaciones</u>	12
<u>Capítulo 2</u>	
<u>Marco Teórico</u>	13
<u>2.1. Definiciones de las Bibliotecas Digitales</u>	14
<u>2.2. Definición de los Repositorios Institucionales</u>	16
<u>2.3. Repositorios de Documentos Administrativos</u>	19
<u>2.4. Archivos Administrativos</u>	21
<u>2.5. Software Libre</u>	21
<u>2.6. El Acceso Abierto</u>	22
<u>2.7. DSpace</u>	24
<u>2.7.1. Capa de Aplicación</u>	26
<u>2.7.2. Capa de Negocios</u>	27
<u>2.7.3. Capa de Almacenamiento</u>	29
<u>2.8. Los Metadatos</u>	29
<u>2.9. Norma ISO 14721</u>	30
<u>2.10. Preservación Digital</u>	32
<u>2.10.1. Metadatos de Preservación</u>	32
<u>2.10.2. PREMIS</u>	33
<u>2.10.3. METS</u>	33
<u>2.10.4. METS-PREMIS</u>	34
<u>2.11. Estándares a Tomar en Cuenta</u>	35
<u>2.11.1. XML</u>	35
<u>2.11.2. OAI-PMH</u>	36
<u>2.11.3. SWORD</u>	37
<u>2.11.4. OpenSearch</u>	37
<u>2.11.5. RSS</u>	37
<u>2.11.6. Search/Retrieve Web service (SRW)</u>	37
<u>2.11.7. Search/Retrieve vía URL (SRU)</u>	37
<u>2.12. Marco Legal en la República Bolivariana de Venezuela</u>	38
<u>2.13. Control de Autoridades de Registros Bibliográficos</u>	39
<u>2.13.1. Clasificación</u>	39
<u>Capítulo 3</u>	

<u>Metodología</u>	<u>40</u>
<u>3.1. Primera Fase: Relevamiento Bibliográfico</u>	<u>40</u>
<u>3.2. Segunda Fase: Análisis e Implementación</u>	<u>40</u>
<u>3.3. Tercera Fase: Documentación Final de la Propuesta</u>	<u>40</u>
<u>Capítulo 4</u>	
<u>Implementación</u>	<u>41</u>
<u>4.1. Planificación</u>	<u>41</u>
<u>4.2. Instalación</u>	<u>41</u>
<u>4.2.1. Entorno de Desarrollo</u>	<u>41</u>
<u>4.2.2. Recomendación One Big Project de DSpace</u>	<u>42</u>
<u>4.2.3. Ant - Maven</u>	<u>43</u>
<u>4.2.4. Recomendación de Proyecto SeDiCI-DSpace</u>	<u>44</u>
<u>4.2.5. Interfaz Gráfica de Usuario: XMLUI</u>	<u>45</u>
<u>4.2.6. Verificación de la Instalación</u>	<u>46</u>
<u>4.3. Configuración</u>	<u>46</u>
<u>4.3.1. Personalización de XMLUI</u>	<u>47</u>
<u>4.3.2. Metadatos, Submission y Workflow</u>	<u>49</u>
<u>4.3.3. Comunidades y Colecciones</u>	<u>52</u>
<u>4.3.4. Control de Autoridades</u>	<u>56</u>
<u>4.3.5. Tipos de Documentos</u>	<u>59</u>
<u>4.3.6. OAI-PMH</u>	<u>63</u>
<u>4.3.7. Discovery</u>	<u>64</u>
<u>4.3.8. Usuarios y Grupos</u>	<u>64</u>
<u>4.4. Plataforma de Trabajo</u>	<u>65</u>
<u>4.4.1. Gestor de Proyectos</u>	<u>65</u>
<u>4.4.2. Entorno de Desarrollo Integrado</u>	<u>65</u>
<u>4.4.3. Servidores</u>	<u>66</u>
<u>4.6. Catalogación</u>	<u>66</u>
<u>4.6.1. Autoarchivo</u>	<u>66</u>
<u>4.6.2. Depósitos por Catalogadores Propios</u>	<u>67</u>
<u>4.6.3. Depósitos Automáticos</u>	<u>67</u>
<u>4.7. Documentación</u>	<u>67</u>
<u>Capítulo 5</u>	<u>68</u>
<u>Consideraciones Finales</u>	<u>68</u>
<u>5.1. Recomendaciones</u>	<u>68</u>
<u>5.2. Conclusiones</u>	<u>68</u>
<u>5.3. Trabajos Futuros</u>	<u>69</u>
<u>Referencias</u>	<u>70</u>
<u>Apéndice A: Informe del Cuestionario</u>	<u>74</u>
<u>Apéndice B: Estructura de la Wiki SeDiCI-DSpace</u>	<u>82</u>
<u>Apéndice C: Siglas y Abreviaturas</u>	<u>83</u>

Índice de Figuras

Figura 1. Repositorios en Venezuela.....	7
Figura 2. Repositorios a Nivel Mundial.....	8
Figura 3. Vías de Publicación.....	22
Figura 4. Arquitectura de 3 Capas - DSpace 1.8.....	25
Figura 5. Modelo Funcional OAIS.....	31
Figura 6. Proveedores del Protocolo OAI-PMH.....	36
Figura 7. Esquema DRI.....	45
Figura 8. Interfaz principal de SeDiCI.....	48
Figura 9. Proceso Ingest.....	50
Figura 10. Proceso de Submission.....	51
Figura 11. Workflows por Defecto de DSpace.....	51
Figura 12. Redmine de SeDiCI.....	65
Figura 13. Entorno de Desarrollo Integrado.....	66

Índice de Tablas

Tabla_1:Paquetes_DSpace.....	26
Tabla_2:DC-PREMIS.....	35
Tabla_3:MODS-PREMIS.....	35
Tabla_4:Formato_XML_Comunidades_Colecciones.....	56

Introducción

El desarrollo de Internet y de las Tecnologías de Información y Comunicación (TIC) han revolucionado el mundo, hecho que muestra inminentemente el cambio de época que se está viviendo hoy en día y que es conocida como la “Era de la Información”. Estas transformaciones son una novedad para la sociedad, la última ocurrió hace aproximadamente más de 200 años cuando la Revolución Industrial condujo a la humanidad del agrarianismo al industrialismo. La génesis del cambio actual se puede observar en lo cultural, social, económico y tecnológico [1 - 2], contando con una sociedad de nativos e inmigrantes digitales [3] que está reestructurando los procesos y la economía mundial, donde definitivamente la materia prima pasó a ser la Información.

En comScore.com [4] se muestran algunas estadísticas importantes que evidencian el uso de la información en esta época:

- Un 65.3% de los usuarios de Internet en el mundo son menores de 40 años.
- Un 17.2% de la población mundial tiene acceso a Internet y en Venezuela es del 36.8%.
- Los sitios más visitados a nivel mundial en estricto orden son: Google, Facebook y Youtube.
- Un 12.2% del comercio electrónico que se realiza es el de libros y revistas, convirtiéndose éstos en los de mayor consumo.
- Sólo el 6.4% de las consultas en Internet son académicas y/o científicas.

Estas estadísticas comprueban el gran auge de la Internet en esta era. El mundo está evolucionando con respecto a la búsqueda y transmisión de la información, así como también en las diferentes concepciones de vida permitiendo soluciones más económicas, un mejor acceso a la información, generación de políticas de conservación del ambiente, etc.

Una posibilidad que concita interés en un grupo importante de personas es el acceso a contenidos y documentos a través de Internet, con las comodidades que esto implica, a la vez que este tipo de acceso posibilita una mejor preselección del material. En el ámbito académico y científico existe un auge de Repositorios Institucionales (RI) a través de comunidades universitarias, principalmente. Este crecimiento de RI se observa en los diferentes directorios de repositorios, por ejemplo, OpenDOAR [5] o ROAR [6].

La producción científica en los últimos años ha crecido en diversas instituciones, a nivel nacional como a nivel internacional, a través de repositorios de producción científica [7]. En Venezuela (ver Figura 1) existen 14 repositorios registrados, para noviembre del 2012, y a nivel mundial (ver Figura 2) se encuentran registrados 2224 repositorios [5]. Los 14 repositorios en Venezuela son:

1. Repositorio de Acceso Abierto EDUMED
2. Scientific Electronic Library Online - Venezuela (SciELO - Venezuela)
3. Saber UCAB
4. Saber UCV

5. Repositorio de producción Intelectual UCLA
6. UCLA - Biblioteca de Administración y Contaduría
7. UCLA Biblioteca de Medicina
8. REVENCYT (Índice y Biblioteca Electrónica de Revistas Venezolanas de Ciencia y Tecnología)
9. Publicaciones realizadas por el Centro Jardín Botánico de Mérida
10. Repositorio Institucional de la Universidad de Los Andes (SABER-ULA)
11. Servicios Bibliotecarios de la Universidad de Los Andes (Serbiula)
12. UDOspace
13. Repositorio Institucional Universidad Nueva Esparta (MiUneSpace)
14. Bolivarium - Universidad Simon Bolivar

De acuerdo con la Figura 2, se evidencia la evolución de esta era de la Información vista desde los repositorios, apoyados en el movimiento mundial de Acceso Abierto. Este movimiento ha terminado de consolidar las herramientas para la preservación y búsqueda de producción académica y de investigación en el dominio de los RI. El crecimiento y desarrollo de los RI en diferentes partes del mundo se debe principalmente a los siguientes beneficios:

- Incorporar a las instituciones de producciones científicas en las comunidades a nivel local y mundial fomentando una cultura de compartir un espacio común a todas las disciplinas.
- Hacer públicas las creaciones científicas de las instituciones en la comunidad local e internacional.
- Realizar búsquedas de material académico de interés a través de diferentes parámetros tales como: búsquedas por autor, título, palabras claves, resúmenes, etc.
- Promover la interoperabilidad entre diversos repositorios a través del protocolo OAI-PMH.
- Acceder a los servicios están disponibles en cualquier momento y desde cualquier lugar las 24 horas al día y 7 días a la semana (24X7).

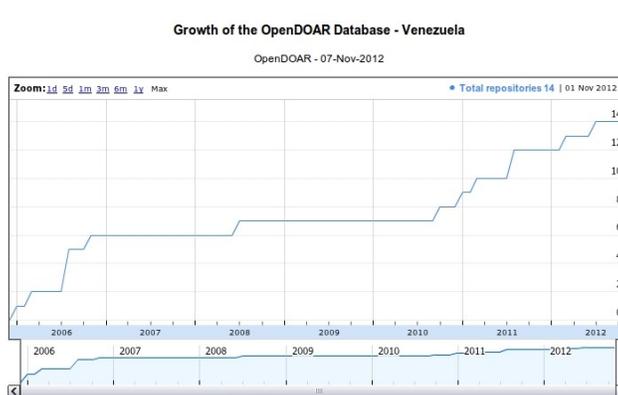


Figura 1. Repositorios en Venezuela. Fuente: OpenDOAR. Nov. 2012.

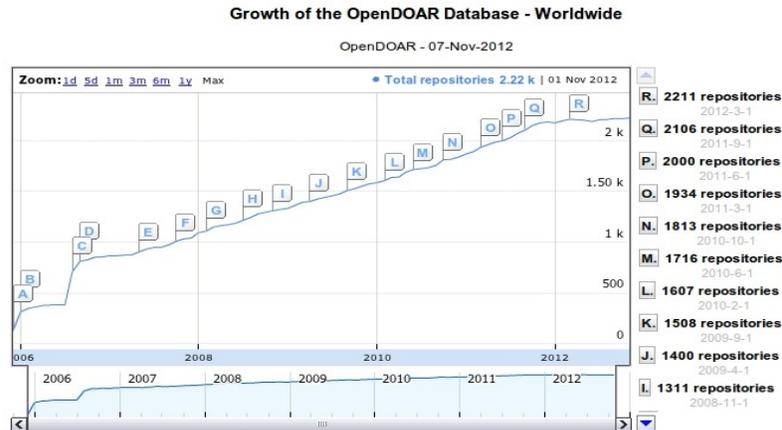


Figura 2. Repositorios a Nivel Mundial. Fuente: OpenDOAR. Nov. 2012.

Los repositorios de documentos digitales administrativos universitarios pasan a ser un espacio ideal para un nuevo ámbito de los repositorios. Por ello, esta propuesta se centra en establecer las bases para implementar un Repositorio de Documentos Administrativos en la Universidad Nacional Experimental del Táchira (en adelante REDAUNET), para permitir la estandarización y fortalecer la gestión de documentos administrativos en instituciones públicas Venezolanas bajo los servicios (recopilación, catalogación, acceso, gestión, difusión y preservación de los documentos) y principios de la filosofía del Software Libre y del Acceso Abierto.

La propuesta está amparada bajo las diferentes leyes y decretos aprobados en la República Bolivariana de Venezuela para la gestión de estos tipos de documentos públicos. La importancia del REDAUNET radica en la necesidad de: (a) establecer y mantener estándares que permitan un crecimiento de ellos, (b) evitar redundancia de trabajo e información para ofrecer un mejor servicio al usuario, (c) reducir los costos y los tiempos en todos los procesos involucrados, y, (d) preservar todos los documentos administrativos.

Los principales beneficios del REDAUNET son:

- Establecimiento de un depósito institucional de documentos administrativos que ofrezca acceso a toda la comunidad universitaria.
- Acceso más fácil y rápido a los documentos para la toma de decisiones en diferentes ámbitos.
- Tener una visibilidad del documento, con el fin de evitar trasladarse a la dependencia de la institución físicamente.
- Acceder a los servicios que estarán disponibles las 24 horas al día y los 365 días del año, desde cualquier lugar y desde diferentes dispositivos de acceso.
- Garantizar una disminución del uso de papel logrando un ahorro económico y beneficios ambientales.

En este trabajo se describen unas consideraciones iniciales para poner en contexto la

propuesta que se presenta, seguido del marco teórico del dominio de los Repositorios Institucionales, después se explica la metodología que se uso para realizar el trabajo, luego se esbozan los pasos para la implementación, y finalmente, se detallan unas consideraciones finales: recomendaciones, conclusiones y trabajos futuros.

Capítulo 1

Consideraciones Iniciales

1.1. Situación Actual

La Universidad Nacional Experimental del Táchira (UNET) es una institución de educación superior pública Venezolana que depende del Consejo Nacional de Universidades a través de la Oficina de Planificación del Sector Universitario del Ministerio del Poder Popular para la Educación Universitaria de la República Bolivariana de Venezuela. La UNET fue creada por Decreto de la Presidencia de la República N° 1630, de fecha 27 de febrero de 1974, con objetivos sobre desarrollo regional señalados en su Reglamento General. El Profesor Lorenzo Monroy Coronel fue el Rector fundador. El Plan de Desarrollo Integral presentado por la Universidad al entonces Ministerio de Educación y al Consejo Nacional de Universidades prevee la conformación de una Universidad que a través de funciones integrales de docencia, investigación y extensión, influya en la formación de recursos humanos, vinculados por el conocimiento de contexto a la solución de los problemas de la región y del país.

La Universidad inició sus labores docentes el 23 de junio de 1975 sobre un proyecto que contempla inicialmente carreras en Ciencias del Agro e Ingenierías de la Industria. En Ciencias del Agro se inició con Ingeniería Agronómica y Zootecnia (actualmente llamada Ingeniería de Producción Animal) y en Ingenierías de la Industria con Ingeniería Industrial a la que después se agregó Ingeniería Mecánica. En 1983 se crea la carrera de Arquitectura. En 1986, gracias a un convenio entre la Universidad del Zulia y la UNET, se inician las actividades de postgrado. En 1991 inicia actividades el programa de Ingeniería Electrónica, en 1997 Ingeniería en Informática e Ingeniería Ambiental en 2004. En 2006 abre la primera carrera humanística de la Universidad, Licenciatura en Música, seguida de seis proyectos de apertura de carreras técnicas cortas en convenio con varias universidades del país.

La UNET tiene 62 unidades administrativas [8] que producen aproximadamente cuatro (4) documentos administrativos diarios en promedio [8], generados en diferentes procesadores de texto, que luego son impresos, entregados y/o guardados en los archivos físicos de cada una de las unidades administrativas. Estas unidades necesitan cada día diferentes documentos relacionados con profesores, trabajadores universitarios, instituciones, estudiantes y/o otras personas en general que tratan temas específicos o procesos administrativos rutinarios, que son requeridos por sus jefes o cualquier otra persona relacionada con la institución.

Los documentos físicos pueden ser perdidos o destruidos fácilmente por alguna actividad natural o actividad humana, por ejemplo, incendios, inundaciones, pérdida de documentos de forma intencionada o no, daños en los discos duros... desencadenantes de un problema institucional producido por la no preservación de los documentos. Por tanto, hacer uso de plataformas informáticas consolidadas (sistemas de repositorios) permite gestionar de una manera fácil y amigable la gestión de todos los documentos generados por cada

dependencia, garantizando su almacenamiento en el tiempo (preservación) a través de la carga de metadatos (características que lo describen) relacionados con el objeto digital.

1.2. Motivación

Existen al menos 2200 sistemas de repositorios implementados de diferentes categorías y tipos, tales como: artículos, datasets, libros, conferencias, software, tesis, objetos de aprendizaje, multimedia, etc. registrados en los directorios de repositorios OpenDOAR y ROAR. Las experiencias de esas implementaciones se relacionan con emular instalaciones realizadas con diversas metodologías para mantener y desarrollar las plataformas de software, con usar las herramientas de desarrollo de software recomendadas y con permitir la interoperabilidad entre los sistemas para supervisar el funcionamiento desde entes centralizados. Por ello, surge la idea de trasladar esta iniciativa a la gestión de documentos administrativos.

Otras de las razones que motivaron esta propuesta fue la aplicación de un cuestionario en la UNET, con el propósito de conocer la gestión y preservación de documentos administrativos (memorandos, resoluciones, oficios, etc.) que se elaboran en las distintas dependencias de la Universidad. El cuestionario (ver Apéndice A) fue aplicado vía Internet de forma anónima a 62 directores o jefes de todas las dependencias administrativas de la Universidad Nacional Experimental del Táchira (UNET). Se completaron 37 cuestionarios (los cuestionarios constan de seis preguntas), estos resultados son interesantes para la investigación porque muestran, grosso modo, cómo y qué se hace con los respaldos de los documentos digitales en las instancias de la universidad, además permitió realizar una propuesta para la gestión y preservación de los documentos administrativos, de tal manera, que se adecuó más a la posición de la Universidad, así como también a la normativa de la UNET y de las leyes del estado Venezolano, ver Apéndice A.

Estas dos razones motivaron el realizar una propuesta a la UNET, y por ende al país, de un REDAUNET que se encuentre bajo el lineamiento del Software Libre, del Acceso Abierto y de la legislación Venezolana que exige, actualmente, el acceso y intercambio de los documentos de las instituciones públicas de todos los Venezolanos.

1.3. Objetivos

1.3.1. General

Diseñar un Repositorio de Documentos Administrativos para la Universidad Nacional Experimental del Táchira (REDAUNET) basado en los documentos generados en todas las dependencias académicas y administrativas de la institución.

1.3.2. Específicos

1. Realizar una revisión teórica de los planteamientos de diversos autores en torno al concepto de los Repositorios de Documentos Administrativos y su importancia de

- acuerdo al marco jurídico de la República Bolivariana de Venezuela.
2. Analizar los tipos de documentos de gestión administrativa en la UNET y hacer una clasificación para adecuarlos al REDAUNET.
 3. Definir el esquema de metadatos de acuerdo con los diferentes tipos de documentos administrativos de la UNET.
 4. Describir el proceso de implementación y mantenimiento de un REDAUNET.
 5. Diseñar documentos de ayuda para la instalación, mantenimiento y desarrollo de la plataforma de software del REDAUNET.

1.4. Alcance y Limitaciones

El alcance está definido en presentar una solución para la gestión y preservación de los documentos administrativos de la UNET, bajo la filosofía del Software Libre y del Acceso Abierto y el cumplimiento de la legislación Venezolana y el reglamento de la UNET. Este trabajo diseñado para una Universidad Venezolana, podría ajustarse a cualquier otra Universidad o institución privada que desee acogerse a los principios de acceso abierto. Para ello, se dispondrá de la documentación y de manuales para realizar los cambios necesarios pero no se explicará detalladamente como desarrollar sistemas de información de este dominio porque se escapan del objetivo principal planteado.

Las limitaciones del trabajo están centradas en la selección de la plataforma para la gestión de repositorios digitales. En este caso, se seleccionó DSpace, cuya plataforma presenta fortalezas (se ajusta a los principios deseados) y debilidades solventadas por desarrollos propios que se escapan al alcance de esta propuesta. La implementación de prueba de la propuesta (accesible vía web) se realizó en los servidores de SeDiCI de la Universidad Nacional de La Plata y el apoyo del personal de esa dependencia, ya que ellos cuentan con la experiencia en esta área. El apoyo está respaldado por el convenio marco de colaboración según resolución número 086/2011 del Consejo Universitario de la UNET.

Por otra parte, esta propuesta será evaluada a través de esta implementación de prueba por las autoridades y jefes de las dependencias de tal institución y en caso de ser necesario, el repositorio deberá ser instalado y migrado a los servidores de la UNET; estas situaciones no están al alcance de nosotros y depende de las autoridades y las dependencias administrativas: CETI (Centro de Estudios en Teleinformática) y Biblioteca de la UNET.

Capítulo 2

Marco Teórico

La concepción que se tiene de bibliotecas tradicionales en las universidades es ofrecer documentos (libros, revistas, tesis, artículos, etc.) en soportes físicos a través de servicios de préstamos y consultas, con el fin de garantizar el apoyo bibliográfico para la docencia, investigación y extensión en la institución de educación. Esas políticas limitaban de alguna manera con el acceso a los documentos por la falta de ejemplares o la existencia de ejemplares desactualizados. No obstante, junto al incremento de los recursos informáticos, la Internet y el descenso de los costos para adquirir esos recursos y servicios relacionados, potenciaron en los últimos 20 años el diseño y la creación de las Bibliotecas Digitales (BD) [9], es decir, se inició la automatización de las bibliotecas tradicionales, con un crecimiento sostenido y en constante evolución. Tales cambios se pueden evidenciar, a partir de varias conferencias internacionales de referencia en el área [10 – 13] entre las cuales se pueden nombrar: Joint Conference on Digital Libraries (JCDL), Theory and Practice of Digital Libraries (TPDL) antes conocida como European Conference on Research and Advanced Technology for Digital Libraries (ECDL), International Conference on Asian Digital Libraries (ICADL), IEEE/TCDL (Technical Committee on Digital Libraries). En los trabajos de estas conferencias y en las publicaciones de revistas del área, 128 revistas en SCImago Journal & Country Rank para mayo del 2012 en la categoría “*Library and Information Sciences*” [14], se evidencia la diversidad de conceptos y descripciones que abarcan las palabras “bibliotecas digitales” o “digital library”. Estos trabajos consolidan esta área en el mundo científico, partiendo del principio que algunas de estas conferencias comenzaron a mediados de los noventa [12 - 13].

Según la bibliografía relevada, a partir de esos años comienzan a surgir las BD, principalmente, gracias a proyectos de investigación financiados por agencias gubernamentales y organismos nacionales e internacionales, mientras que otros específicos a cargo de instituciones académicas, de investigación y de bibliotecas, individualmente o en colaboración. En 1993 a través de 6 proyectos, de las Universidades de Michigan, Stanford, Berkeley, Santa Barbara, Illinois y Carnegie Mellon [15], en Estados Unidos de América surge Digital Library Initiative (DLI-1) integrada por National Science Foundation (NSF), Defense Advanced Research Projects Agency (DARPA) y National Aeronautics and Space Administration (NASA). El objetivo de DLI-1 fue desarrollar e implementar modelos de bibliotecas digitales, logrando avanzar considerablemente en poder recopilar, almacenar y hacer disponible para la búsqueda, recuperación y procesamiento, documentos científicos y de investigación a través de las redes de comunicación. Luego de los logros obtenidos por la DLI-1, en 1998 se anuncia un nuevo programa que fue llamado Digital Libraries Initiative - Phase 2 (DLI-2), conformado por National Library of Medicine (NLM), the Library of Congress (LC), the Federal Bureau of Investigation (FBI) y the National Endowment for the Humanities (NEH), además de los organismos que llevaron a cabo la DLI-1 [16]. Los resultados de estas dos iniciativas permitieron la consolidación y estudio de nuevos estándares aplicados hoy en día en las Bibliotecas Digitales.

2.1. Definiciones de las Bibliotecas Digitales

Un primer concepto se puede encontrar en 1997, cuando el autor Lesk [17] expresa: “Las bibliotecas digitales son colecciones organizadas de información digital. Combinan la estructura y concurrencia de la información, que siempre han tenido las bibliotecas y los archivos, con la representación digital que han hecho posible las computadoras”.

En 1998 Waters [18], expuso una definición de las bibliotecas digitales para que todas las instituciones en la Digital Library Federation compartieran un mismo concepto, el cual podría someterse a revisiones. La misma consistió explícitamente en definir las como organizaciones que proveen los recursos tecnológicos para seleccionar, estructurar, controlar el acceso, conservar la integridad y asegurar la persistencia, a través del tiempo, de colecciones de trabajos digitales que estén disponibles para usarse por una o varias comunidad definidas.

En 1999, Borgman [19] establece que las bibliotecas digitales son un conjunto de recursos electrónicos y de capacidades técnicas asociadas para crear, buscar y utilizar la información, es decir, son una extensión y mejora de sistemas de almacenamiento y recuperación que manipulan los datos digitales en cualquier medio. También señala la existencia de dos visiones diferentes, una referida a las comunidades de investigadores y la otra referida a los profesionales de las bibliotecas. Este autor dice explícitamente lo siguiente: “... desde la perspectiva de investigación, las bibliotecas digitales son contenido recopilado y organizado en virtud de comunidades de usuarios ... desde la perspectiva de la práctica bibliotecaria, las bibliotecas digitales son instituciones u organizaciones que proveen servicios de información en formas digitales”.

Chowdhury y Chowdhury [11] en 1999 exponen un panorama general sobre el área y se pone en evidencia la proliferación de trabajos de investigación en los diferentes procesos presentes en las bibliotecas digitales, así como también la convergencia de bibliotecarios, informáticos, ingenieros electrónicos, abogados, científicos cognitivos, etc. En resumen, estos autores indican que las BD están todavía en una fase experimental y que se debe esperar un tiempo para terminar de ver totalmente implementada una BD.

También en 1999, Bawden y Rowlands [20] exponen en su trabajo que las BD utilizan las expresiones biblioteca electrónica, virtual y digital. Estas expresiones en muchas ocasiones trabajan como sinónimas en la bibliografía especializada. Por tanto, para estos autores una biblioteca electrónica estaría formada por objetos físicos que necesitan de medios electrónicos para el acceso a la información contenida en los mismos. La biblioteca virtual sería aquella que hace uso de la realidad virtual para mostrar una interfaz y emular un ambiente que sitúe al usuario dentro de una biblioteca tradicional. Y la biblioteca digital consistiría en un conjunto de materiales y servicios almacenados, procesados y accedidos es un repositorio de colecciones y contenidos digitalizados, almacenados en diferentes formatos electrónicos, procesados y accedidos mediante la utilización de herramientas y redes de comunicaciones digitales.

Kochtanek y Hein en 1999 [21] analizan la percepción de los investigadores y usuarios sobre varios aspectos de la biblioteca digital, incluyendo su conceptualización, a través del método Delphi. Las principales conclusiones fueron que las iniciativas existentes hasta el momento son principalmente de colaboración, la biblioteca digital tiene el potencial para transformar el acceso a los registros digitales de conocimiento y que el principal rol de los bibliotecarios es dar una extensión más al mundo de las bibliotecas digitales.

En el 2000, Arms [22] propuso una definición informal: “una biblioteca digital es una colección gestionada de información, con servicios asociados, donde la información es almacenada en formato digital y es accesible en toda la red”. Esta definición enfatiza los aspectos de la gestión de los contenidos. Ese mismo año, Sharon y Frank [23] presentaron un trabajo al IFLA Annual Conference en Jerusalem, donde definieron a la biblioteca digital como una biblioteca computarizada en donde la mayor parte de la información es digital y definieron una clasificación en 3 categorías: *Stand-alone Digital Library*: que consiste en una biblioteca simple en donde los recursos son digitales, *Federated Digital Library* es una federación de varias *Stand-alone Digital Library* independientes en la red, organizadas en un tema en común, y, *Harvested Digital Library* es una biblioteca virtual que proporciona acceso a material diseminado en la red.

En el 2002, Jesus Tramullas [24] afirmó que las BD llegaron a un nivel de madurez, tal que, el objeto de estudio estuvo lo suficientemente definido por sus propias funciones y herramientas así como por los componentes tecnológicos, y define a la BD como: “un sistema de tratamiento técnico, acceso y transferencia de información digital, estructurado alrededor del ciclo de vida de una colección de documentos digitales, sobre los cuales se ofrecen servicios interactivos de valor añadido para el usuario final.” De igual manera, comenta que más allá de que existan las bibliotecas “híbridas”, la biblioteca digital emerge independientemente, en parte por el impulso enorme de un cuerpo de investigadores de la mitad a finales de la década de los noventa y por el apoyo de las agencias gubernamentales.

En el 2003, Ian Witten et al. [25], afirman que las BD son colecciones enfocadas de objetos digitales, incluyendo texto, vídeo y audio, así como los métodos de acceso y recuperación de datos, y para la selección, organización y mantenimiento de la colección. Tradicionalmente estos se dividen en búsqueda y consulta, aunque en realidad la distinción no es clara.

Luego, Tramullas en el 2007 [26] comenta que no puede considerarse a las BD como bases de datos ni una digitalización de textos en la internet, tampoco a un OPAC (Online Public Access Catalog) integrado con enlaces y/o una colección de documentos en PDF o en cualquier otro formato. Por ello, define tres tipos principales de expresiones bibliotecarias: la biblioteca automatizada que es aquella en la cual los procesos clásicos de tratamiento y recuperación son desarrollados mediante la utilización de sistemas informáticos, la biblioteca digitalizada que es aquella en la cual se han desarrollado procesos de digitalización de fondos por diferentes razones y finalmente la biblioteca digital como aquella que pone a disposición de los usuarios una colección digital, sobre la cual se

construyen servicios de valor añadido.

En el 2007, Candela et al. publican el manifiesto de DELOS [27], donde establecen un marco conceptual con tres niveles:

- Biblioteca digital es la organización que recopila, gestiona, preserva y ofrece contenidos digitales,
- Sistema de biblioteca digital que se refiere al sistema de software que proporciona la funcionalidad requerida por una biblioteca digital particular y,
- Sistema de gestión de biblioteca digital que se refiere a la plataforma: sistema operativo, bases de datos, interfaz de usuario.

En el 2010, Linghui Guo [28] define a las BD como aquellas que se procesan con tecnología digital y permite el almacenamiento de diversos documentos, además de presentarse como un sistema de información distribuido con el almacenamiento de los recursos de información de diferentes fuentes y de diferentes ubicaciones geográficas. Las características más importantes de las bibliotecas digitales que este autor expone son la digitalización de la colección, operaciones computacionales, transmisiones a través de la red, apertura del resguardo de la información y el compartir los recursos.

En el 2011, Nguyen y Chowdhury [10] realizan un estudio sobre los tópicos y subtópicos de las BD que permiten crear un plan de estudios para un curso sobre el área, convirtiéndose en una mapa conceptual ideal para investigadores de las BD, profesores, estudiantes y profesionales.

Luego de leer todas estas definiciones presentadas en diferentes trabajos desde 1993 hasta el presente, se observa que el concepto de Bibliotecas Digitales ha ido evolucionando en todos esos años; la expresión “bibliotecas digitales” es usada como un gran término, dentro de la cual se engloban actividades de investigación y de desarrollo muy variadas, con diferentes objetivos, y bajo diferentes perspectivas e interpretaciones.

En conclusión una biblioteca digital es un sistema de información que permite el acceso y transferencia de información digital, estructurada alrededor de colecciones de documentos digitales sobre los cuales se ofrecen servicios a los usuarios. Este concepto no es definitivo, porque se ha observado como han evolucionado y seguirán creciendo en paralelo al crecimiento de la tecnología, representada por la Internet, el software y/o el hardware.

2.2. Definición de los Repositorios Institucionales

Hoy en día, se confunden y aúnan las definiciones de las Bibliotecas Digitales con las definiciones de los Repositorios Institucionales (RI), de hecho los RI comparten muchas características con las BD, pero se pueden distinguir algunas características importantes. Los repositorios, también conocidos como repositorios digitales, están constituidos por un conjunto de archivos digitales en representación de productos científicos y académicos que pueden ser accedidos por los usuarios. Específicamente, los RI consisten en estructuras web

interoperables de servicios informáticos, dedicadas a difundir la perpetuidad de los recursos científicos y académicos (físicos o digitales) de las universidades a partir de la enumeración de un conjunto de datos específicos (metadatos), para que esos recursos se puedan recopilar, catalogar, acceder, gestionar, difundir y preservar, por lo general están estrechamente ligados a los ideales y objetivos del Acceso Abierto. La representación de estos recursos se logra mediante el registro persistente del conjunto de datos asociados a ellos. Estos datos sirven como síntesis y reemplazo del objeto "real", lo cual permite distribuir el recurso sin requerir del objeto en sí, sino usando su representación. Las actividades de catalogación, acceso, gestión y difusión de los contenidos son las más consolidadas con el crecimiento de los repositorios, por el contrario, la recopilación de materiales y la preservación todavía se encuentran en sus primeros pasos. En fin, se pueden enumerar las siguientes características:

- Los RI pueden estar agrupados en directorios de repositorios. Los directorios más referenciados de acuerdo con los enlaces entrantes o inlinks, según los buscadores web Majestic SEO y ahrefs, datos tomados el 1 de noviembre del 2012, son: -OpenDOAR tiene registrados 2217 repositorios registrados [5]. -ROAR - Registry of Open Access Repositories tiene registrados 2963 repositorios registrados [6]. -University of Illinois OAI-PMH Data Provider Registry, con 2946 repositorios registrados [29].
- La producción intelectual de las instituciones está presente en el RI en una colección de documentos y objetos, de tipología y formatos diversos.
- El depósito de los textos, conjuntos de datos, archivos de sonido, imágenes o cualquier otro tipo de documento puede ser realizado directamente por el investigador/institución afiliado mediante el autoarchivo o desde la administración.
- Los documentos que se depositan pueden estar en cualquier etapa del proceso de la investigación académica, esto depende de la política de la institución sobre los documentos que se pretenden depositar.
- Muchos RI están relacionados con los ideales y objetivos del Acceso Abierto (Open Access).

Uno de los primeros autores que introduce el concepto de RI es Crow en el 2002 [30], indicando que son colecciones digitales que capturan y preservan la producción intelectual de las comunidades universitarias, pudiendo ampliar su campo de implementación.

Clifford Lynch [31] en el 2003, señala: “un repositorio institucional universitario es un conjunto de servicios que ofrece la universidad a los miembros de su comunidad para la dirección y distribución de materiales digitales creados por la institución y los miembros de esa comunidad. Es esencial un compromiso organizativo para la administración de estos materiales digitales, incluyendo la preservación a largo plazo cuando sea necesario, así como la organización, acceso y/o distribución”.

En el 2004, Van de Sompel et al. [32] expresa que “un sistema de comunicación académica futura que incorpore de forma nativa la capacidad de grabar y exponer esa dinámica, las relaciones y las interacciones en la propia infraestructura de comunicación académica. El

registro de este cuerpo de información es sinónimo de registrar la evolución de la academia en una granularidad fina. Esto permitirá rastrear el origen de las ideas específicas hasta sus raíces, analizando las tendencias en un momento específico en el tiempo y proyectando las futuras líneas de investigación”.

Tramulas y Garrido en el 2006 expresan [33] que los RI se han convertido en la principal forma de publicar, preservar y difundir la información digital de las instituciones, gracias a la conjunción del acceso abierto, del software libre y de los estándares abiertos aplicados en este dominio. También señalan que eso se añade al desarrollo de las revistas y publicaciones electrónicas de calidad, que permite a las comunidades investigadoras estar actualizadas y comunicar los avances obtenidos.

Luego la Universidad Autónoma de Madrid [34], en el 2007, define a un RI como un conjunto de servicios web centralizados, creados para organizar, gestionar, preservar y ofrecer acceso libre a la producción científica, académica o de cualquier otra naturaleza cultural, en soporte digital, generada por los miembros de una institución.

Bustos y Fernández en el 2008 [35] publican unas directrices para la creación de RI en universidades y organizaciones de educación superior, y definen a los RI como "un archivo electrónico de la producción científica de una institución, almacenada en un formato digital, asimismo contiene mecanismos para importar, identificar, almacenar, preservar, recuperar y exportar un conjunto de objetos digitales, normalmente desde un portal web". Afirman que el RI es una vía de comunicación científica, pero no puede ser entendido como un canal de publicación, sino que debe comprenderse como un complemento al proceso de publicación científica formalizado con revisión por pares.

Luego de leer y comprender algunos conceptos de las BD (sección anterior) y de los RI, se presentan a continuación algunas diferencias entre ellos:

- Los RI están organizados en relación con una comunidad institucional en particular, en cambio las BD están construidas en torno a un número diferente de posibles principios organizativos: tópico, sujeto, disciplina, etc.
- A menudo los RI dependen de la contribución voluntaria conocida como autoarchivo. Las BD son producto de una estrategia deliberada de desarrollo de colecciones por parte de los profesionales de la biblioteca, donde muchas veces tienen contenidos más allá de la propiedad institucional.
- Los RI son principalmente repositorios y pueden ofrecer servicios limitados a los usuarios. Las BD incluyen aspectos importantes de servicio: referencia, asistencia, interpretación de contenidos, es decir, apoyo de personal en la búsqueda de información adicional.
- Los RI tienen que ser interoperables a través de estándares para la colecta de la información que administran, en cambio, las BD tienen sus propios sistemas y no se preocupan por ser interoperables con otros sistemas.
- Los elementos de un repositorio suelen ser de origen digital, las bibliotecas digitales por lo general tienen que digitalizar los viejos materiales (analógico), además de los

de origen digital que pueden tener.

Es necesario aclarar que en este trabajo se utilizan indistintamente los términos RI y BD gracias a la evolución de los servicios e infraestructuras web, puesto que, según algunos autores, un Repositorio Institucional es una Biblioteca Digital y una Biblioteca Digital es un Repositorio Institucional [36].

2.3. Repositorios de Documentos Administrativos

Los repositorios de documentos digitales se están convirtiendo progresivamente en los lugares en los que numerosas organizaciones almacenan y organizan el resultado de sus actividades académicas, científicas y administrativas. Después de relevar información sobre BD y RI se replican para la consolidación de un Repositorio de Documentos Administrativos (RDA), que permita a una organización o institución gestionar la información administrativa de cualquier naturaleza en forma electrónica y almacenarla en un formato determinado, de esa manera los RDA podrán ofrecer los servicios de recopilación, catalogación, acceso, gestión, difusión y preservación de la misma forma que los RI.

Además de los RDA se encuentran los Gestores Documentales (GD) que son herramientas que facilitan la creación, edición y compartición de documentos electrónicos generados por unidades o dependencias organizativas que permiten a los usuarios leer, actualizar y/o borrar esos documentos, es decir, manejar los documentos en línea. A continuación se enumeran unas diferencia entre los RDA y los GD:

- **Tipos de documentos:**
RDA: documentos de procesadores de texto.
GD: documentos en producción y trámite.
- **Modificación de documentos:**
RDA: prohibido, solo se podrá incluir la versión final.
GD: está permitido, es decir, pueden existir varias versiones de un mismo documento o un historial.
- **Versiones de documentos:**
RDA: solo se conserva la versión final y no existirán versiones de los documentos.
GD: pueden conservarse y administrarse diferentes versiones de los documentos.
- **Eliminación de documentos:**
RDA: prohibido, a menos que se quiera eliminar el recurso completo del RDA.
GD: permitido, pero queda en el histórico.
- **Sistema de clasificación y almacenamiento:**
RDA: obligatorio, ya que los catalogadores necesitan hacer una clasificación de cada uno de ellos.
GD: opcional, ya que suelen depender del usuario.
- **Políticas de preservación:**
RDA: los sistemas garantizan una conservación a largo plazo.
GD: opcional.

- **Objetivo:**
RDA: proporcionar un repositorio seguro para la preservación y el acceso a los recursos.
GD: facilitar la gestión de documentos en el trabajo diario.

Una vez analizados los conceptos de los RDA y los GD queda claro que el alcance y el objetivo principal de la propuesta es un Repositorio de Documentos Administrativos para la UNET y no una gestión de ellos, aunque existe la posibilidad de incluir módulos que permitan ampliar las funciones del repositorio.

Existen cuatro normas importantes que han contribuido y pueden contribuir para desarrollos futuros de este dominio:

- Norma ISO 14781 conocida como *Reference Model for an Open Archival Information System* (OAIS), modela las partes que componen un sistema abierto de archivo de información sin definir su implementación (en la sección 2.8 de este trabajo se profundizará sobre la descripción de esta norma).
- Norma ISO 23081, ofrece una recomendación para los procesos de gestión de documentos.
- Norma ISO 15489 y ISO 26122, que también se refieren indirectamente a las necesidades de los procesos de negocio y los requisitos de los documentos.

Los trabajos de Devakos y Toth-Waddell [37], en el 2008, describen un proyecto para mejorar el acceso y la longevidad de los documentos del gobierno electrónico usando DSpace y concluyen que este software puede ser usado para dicho objetivo. La longevidad depende directamente del medio de almacenamiento usado y del formato privativo o libre del documento. En el 2009, CENATIC (Centro Nacional de Referencia de Aplicación de las TIC basadas en fuentes abiertas) [38] desarrolla un trabajo donde exponen una descripción de un archivo electrónico de las Administraciones Públicas en España y hace un análisis de las plataformas de software abiertas existentes.

En el 2011 se publican tres trabajos que fortalecen la consolidación de repositorios para las administraciones públicas. El trabajo de Moreira et al. [39] realiza una evaluación de tres aplicaciones de software libre para la gestión de archivos administrativos para su implementación en una universidad y los resultados indicaron que las aplicaciones analizadas no cumplen con todos los requisitos de un gestor de archivos, pero por ser desarrollo en software libre puede permitir actualizaciones e integraciones para cumplir con dichos requisitos. El estudio de Rodríguez-Vela [40] parte de cómo las administraciones públicas son generadoras de información y por tanto, pueden difundir, conservar y reutilizar esa información para llevar a cabo la misión de servicio público que tiene encomendada, evidenciando la necesidad de contar con sistemas que permitan la preservación y el acceso para la comunidad. El tercer trabajo es desarrollado por Marcos-Martin y Soriano-Maldonado [41] que manifiestan la importancia de la reutilización de la información del sector público como un elemento clave en las estrategias de *open data* y *open government*, dinamizando la economía digital que actúa como pilar de la transparencia democrática.

2.4. Archivos Administrativos

Los archivos son depósitos donde se almacenan en forma organizada los documentos con el fin de garantizar los servicios administrativos o de servir de fuentes para investigaciones [42]. La tipología archivística puede ser clasificada según la entidad productora en públicos y privados, según su finalidad en administrativos e históricos, y, según su norma clasificatoria en estatales, municipales, eclesiásticos y particulares.

En el ámbito de este trabajo, se limitará a tomar en cuenta los documentos que pertenecen en cierto modo a los archivos administrativos porque estos documentos, por lo general, no se encuentran en el archivo institucional de la UNET, a pesar que los archivos que hoy son históricos, anteriormente fueron administrativos [42].

2.5. Software Libre

El software libre es todo producto de software que puede ser distribuido, modificado, copiado y usado, y debe acompañarse del código fuente para hacer efectivas las cuatro libertades que lo caracterizan:

1. Libertad 0: la libertad para ejecutar el programa sea cuál sea el propósito.
2. Libertad 1: la libertad para estudiar el funcionamiento del programa y adaptarlo a las necesidades propias.
3. Libertad 2: la libertad para redistribuir copias.
4. Libertad 3: la libertad para mejorar el programa y luego distribuirlo para el bien de toda la comunidad.

Tal y como lo afirma Richard Stallman [43], fundador del movimiento y de *Free Software Foundation* (FSF), el software libre es una cuestión de libertad no de precio, es decir, se debe pensar en la acepción de libre como en libertad de expresión y no gratuidad. En conclusión, el software libre refiere a la libertad de los usuarios para ejecutar, copiar, distribuir, estudiar, cambiar, mejorar el software y poder distribuir lo modificado.

A Stallman le pareció poco viable, a corto plazo, eliminar las leyes del copyright así como las injusticias que consideraba provocadas por el marco legal del copyright al no permitir las cuatro libertades sobre los software desarrollados, y creó su propia licencia de derechos de autor, Licencia Pública General de GNU (GPL), la cual es la forma más simple de hacer que un programa sea libre y de dominio público, sin derechos reservados, es decir, sin las leyes del copyright. A partir de este marco legal surge el copyleft que es la forma de ejercer el derecho de autor permitiendo la libre distribución de copias y versiones modificadas de una obra u otro trabajo, además de exigir que los mismos derechos sean preservados en las versiones modificadas.

A partir de la iniciativa del software libre surgió el movimiento del código abierto (*open*

source) que se entiende como una filosofía que promueve la libre distribución y el acceso al diseño de un producto final de software. Este nació de la mano de algunos usuarios de la comunidad del software libre, que se usa como reemplazo al ambiguo nombre original en inglés (*free software*). Aunque en la práctica el software de código abierto y el software libre comparten muchas de sus licencias, la FSF opina que son movimientos filosóficamente diferentes, ya el software libre hace especial énfasis en los aspectos morales o éticos, en el que la excelencia técnica es vista como un producto secundario. Por el contrario, el código abierto busca darle mayor relevancia a los beneficios prácticos de compartir el código fuente e interesar a las principales casas de software y otras empresas de la industria de la alta tecnología, por lo cual hace de la excelencia técnica el objetivo prioritario, en otras palabras, compartir el código fuente es un medio para dicho fin.

2.6. El Acceso Abierto

El Acceso Abierto, también conocido como *Open Access* (OA), tiene como fin asegurar el acceso inmediato, sin restricciones y libre a la producción científica, académica o de cualquier otro tipo, es decir, el usuario puede leer, descargar, copiar, distribuir, imprimir, buscar o enlazar los textos de diversas producciones. El movimiento de acceso abierto a la información se basa en dos estrategias fundamentales, una a través de las revistas de acceso abierto y la otra por medio de los repositorios institucionales. En 1966, se inicia el movimiento con el lanzamiento de *Educational Resources Information Center* (ERIC), biblioteca digital especializada en educación, y con Medline, una base de datos bibliográfica de biomedicina producida por la *National Library Medicine* (NLM) de los Estados Unidos. Una de las voces líderes es Peter Suber [44] que en el 2005 indicó que existe una gran división en las publicaciones científicas, una referida a aquellas que están disponibles gratuitamente en la Internet (ruta verde) y otra en las cuales los autores deben pagar para que el público tenga acceso a ellas (ruta dorada) (ver Figura 3). El autor enumera entre los beneficios del acceso abierto que los artículos son citados 50-300% más que aquellos que no están en OA en una misma revista y resalta la importancia del autoarchivo como bandera del movimiento.

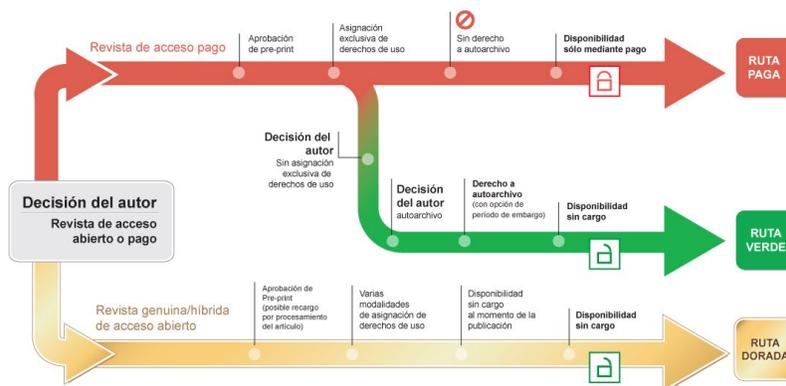


Figura 3. Vías de Publicación. Fuente: SeDiCI basado en la fundación Max Planck

Otro de los trabajos importantes de Suber es el *timeline* publicado sobre el OA a través de su página web [45]. A continuación se resaltan los puntos más importantes a nuestro entender del *timeline* de Suber y de *Open Access Directory* [46]:

En 1966:

- El lanzamiento de ERIC.
- El lanzamiento de Medline (disponible en la red en 1997). Los inicios del movimiento por el acceso libre a la literatura científica revisada por pares se encarnan con la historia misma de Internet.

En 1971:

- *Project Gutenberg* creado por Michael Hart.

En 1991:

- Lanzamiento por Paul Ginsparg en agosto de 1991 de ArXiv.

En 1994:

- Stevan Harnad, uno de los líderes más activos del movimiento lanza la iniciativa del autoarchivo.
- *National Science Foundation* y otras agencias de los Estados Unidos crean el proyecto *Digital Libraries Initiative*.
- Proyecto del Genoma Humano lanza su sitio web en acceso abierto.

En 1998:

- Declaración de San José.

En 1999:

- Se crea el *Open Archives Initiative* (OAI).
- Convención de Santa Fe en Albuquerque, Estados Unidos.

En 2000:

- Pubmed Central.
- *Declaration on Science and the Use of Scientific Knowledge* por parte de la UNESCO.
- BiomedCentral.
- *Public Library of Science* (PloS) .
- Universidad de Southampton lanza Eprints.

En 2002:

- Budapest *Open Access Initiative*.
- *Creative Commons* por Lawrence Lessig.
- IFLA publica su manifiesto.
- MIT lanza el DSpace.

- *Public Knowledge Project* crea el *Open Journal Systems (OJS)*.

En 2003:

- *Bethesda Statement on Open Access Publishing*.
- *Berlin Open Access Initiative (BOAI)*.
- La Universidad de Virginia y Cornell crean FEDORA.

En 2008:

- Peter Suber dejó de actualizar su lista de hechos del OA y la lista se movió a *Open Access Directory*.
- Se creó el día del *Open Access*, 14 de octubre.

En 2009

- Se creó la semana del *Open Access*, que se ha celebrado en el 2010, 2011 y 2012 en diferentes países.

Han surgido otros movimientos a partir del *Open Access* como el *Open Data* y el *Open Government*. El primero tiene una ética similar al software libre, el código abierto (*open source*), el acceso libre (*open access*) y el contenido abierto (*open content*). Todos estos movimientos abiertos aconsejan el uso de los modelos de licencia *Creative Commons*, inspirados en las GNU GPL. Esta licencia se constituyó en 2002 por algunos profesores y profesionales norteamericanos y británicos, liderada por el profesor Lawrence Lessig de la Universidad de Stanford.

El *Open Government* es la doctrina política que sostiene que los temas de gobierno y administración pública deben ser abiertos a todos los niveles posibles en cuanto a transparencia. Se basa en tres pilares: la transparencia en la gestión pública, el fomento de la participación y la colaboración de la sociedad en el diseño y ejecución de las políticas públicas. Por tanto, aboga porque las administraciones públicas pongan a disposición de la sociedad la información pública que tienen en su poder.

2.7. DSpace

DSpace es una herramienta de código abierto desarrollada por el Instituto Tecnológico de Massachusetts (MIT) en colaboración con HP para la implementación de su propio repositorio. DSpace fue liberada en el 2002, se presenta como una solución completa y proporciona toda la funcionalidad necesaria de un repositorio digital que permite la administración de colecciones digitales tales como libros, artículos, fotos, videos, tesis y otros. Los datos son organizados como ítems que pertenecen a una colección y esta a su vez pertenece a una comunidad. Según OpenDOAR [5] están registrados 2217 repositorios y el 41.23% (904) usa DSpace, lo que lo convierte el tipo de software líder para repositorios.

Esta herramienta está constituida actualmente sobre una arquitectura de tres capas, formada por módulos y componentes independientes, lo que favorece un encapsulamiento de la

funcionalidad. El *core* está conformado por los módulos ubicados en la *capa de negocio*, quien negocia la gestión de los contenidos de los archivos junto con sus usuarios, autorizaciones y flujos de trabajos (*workflows*). La inclusión de funciones en la *capa de aplicación*, se hace gracias a add-ons. La comunicación entre los distintos componentes de la aplicación se realiza a través de componentes que comunican el mundo exterior y el DSpace, para ello usan la API pública de DSpace. Finalmente la *capa de almacenamiento*, usa una base de datos relacional para almacenar todo el contenido de la organización, los metadatos de los contenidos, la información sobre los usuarios de los archivos, las autorizaciones y las diferentes etapas de los *workflows* que se están ejecutando. Al igual que la capa anterior, se cuenta con una API para esta capa.

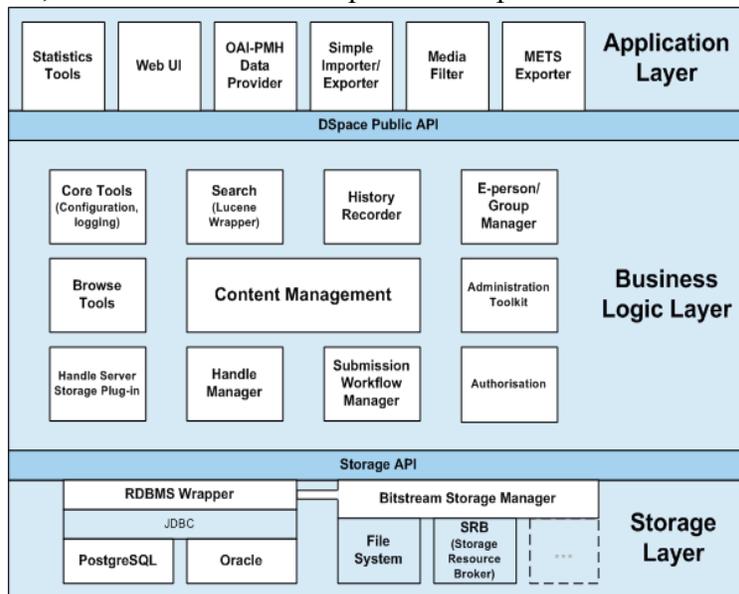


Figura 4. Arquitectura de 3 Capas - DSpace 1.8

Como se observa en la figura anterior, cada capa es invocada por la capa que se encuentra debajo de ella, por ejemplo, la capa de aplicación no puede ser usada de forma directa por la capa de almacenamiento. También se observan las APIs de las capas de almacenamiento y negocios llamadas Storage API y DSpace Public API respectivamente. Estas APIs son métodos, objetos y clases de Java.

Aunque la lógica para las acciones de autorización están en la capa de negocio, el sistema se basa en las aplicaciones individuales en la capa de aplicación para una autenticación correcta y segura de los usuarios de archivos. La razón de esta elección de diseño es que los métodos de autenticación pueden variar entre las diferentes aplicaciones, así que tiene sentido dejar la lógica y la responsabilidad en esas aplicaciones.

El código fuente está organizado para respetar en forma estricta la arquitectura de tres capas y sólo a través de los métodos de la API pública de un componente se les da acceso. Esto significa que Java ayuda a asegurar que el código fuente se ajuste a la arquitectura. Por ello, los componentes de cada capa están distribuidos por paquetes como se muestra en la

siguiente tabla:

Capa	Paquetes de Java
Aplicación	org.dspace.app
Negocios (no incluye app ni storage)	org.dspace
Almacenamiento	org.dspace.storage

Tabla 1. Paquetes por cada capa de DSpace

En conclusión, cada capa está estructurada de la siguiente manera y será explicada en detalle:

1. Capa de Aplicación (*Application Layer*)
 - *Web User Interface*
 - *OAI-PMH Data Provider*
 - *Item Importer and Exporter*
 - *Transferring items Between DSpace Instances*
 - *Registration*
 - *METS Tools*
 - *Media Filters*
 - *Sub-Community Management*
2. Capa de Negocios (*Business Logic Layer*)
 - *Core Classes*
 - *Content Management API*
 - *Workflow System*
 - *Administration Toolkit*
 - *E-person/Group Manager*
 - *Authorisation*
 - *Handle Manager/Handle Plugin*
 - *Search*
 - *Browse API*
 - *History Recorder*
 - *Checksum Checker*
3. Capa de Almacenamiento (*Storage Layer*)
 - *RDBMS*
 - *Bitstream Store*

2.7.1. Capa de Aplicación

2.7.1.1. *Web User Interface*: este componente conocido como WebUI es el más utilizado en DSpace, ya que permite que los usuarios accedan a DSpace a través de *browsers* o navegadores. Actualmente, están disponibles dos proyectos WebUI basados en sistemas Maven, JSPUI y XMLUI (cada uno separado) y usan Ant para desplegar sus *webapps*. JSPUI trabaja con un modelo similar al MVC donde la vista son JSPs, los controladores son Java Servlets y los modelos contenidos API. El proyecto JSPUI con el pasar del tiempo ha

dejado de ser popular para dar paso a XMLUI. XMLUI también conocida como proyecto Manakin está desarrollada a partir del *framework Cocoon* y dividida en tres niveles: *Style Tier* (HTML+CSS), *Theme Tier* (XSL+XML+CSS) y *Aspect Tier* (JAVA+COCOON). Cocoon está diseñado alrededor del concepto de SoC (*Separation of Concerns*) y usa una arquitectura basada en componentes. Se concentra en la publicación de *websites* utilizando SAX (Simple API for XML) poniendo fuerte concentración en los *pipeline*. El principio de SoC es permitir el desarrollo paralelo para que cada desarrollador se enfoque en su parte sin afectar a los otros. Los desarrolladores producen operaciones (componentes) que se unen para formar un *pipelines*. Algunos de los tipos de componentes existentes son: *matchers*, *generators*, *transformers*, *serializer*, *selectors*, *views*, *readers* y *actions*. Estos componentes no están interconectados; tampoco hay llamadas a métodos entre ellos. En cambio su interacción es guiada por su constructor que en la mayoría de los casos es el *sitemap*, definido como un conjunto de documentos XML que describen como todos los componentes de Cocoon son configurados juntos.

2.7.1.2. *OAI-PMH Data Provider*: DSpace soporta el *Open Archive Initiative Protocol for Metadata Harvesting* (OAI-PMH) en su versión 2.0 como un proveedor de datos. Este protocolo establece un conjunto de reglas a partir de las cuales se puede realizar un intercambio de recursos de forma confiable y exitosa, este proceso también es conocido como interoperabilidad. Este servicio es desplegado como otra aplicación web de DSpace, por ello es independiente de los WebUI y se ubica en la capa de aplicación. Este servicio puede ser usado a través del URL de la instalación realizada.

2.7.1.3. *DSpace Command Launcher*: se permite que comandos y *scripts* puedan ser ejecutados vía consola o CLI (*Command Line Interface*).

2.7.2. Capa de Negocios

2.7.2.1. *Core Classes*: a través del package `org.dspace.core` se ofrecen un conjunto de clases básicas que pueden ser utilizadas por los diferentes componentes de DSpace. Una breve explicación de cada una de ellas:

El *Configuration Manager* es la clase responsable de leer el principal archivo de configuración del sistema (`dspace.cfg`), de administrar los archivos de configuración de aplicaciones como Apache y de obtener los textos de los correos electrónicos.

La clase *Constants* es la que administra las constantes que son usadas para representar los tipos de objetos y acciones en la base de datos.

La clase *Context* es la principal para la operación de DSpace. Cualquier clase que desee utilizar la API de la capa de negocio, debe primero crear un objeto de esta clase, es decir muy similar a la apertura de una base de datos. Esta clase está involucrada en muchas llamadas de métodos y constructores de objetos, entonces esos métodos y objetos tienen acceso a información acerca de la operación actual. Por tanto, cuando un objeto *Context* es creado, automáticamente se inicia una conexión a la base de datos, un caché de los objetos API del *configuration manager* y una sesión de usuario autenticada junto con toda la información relacionada del usuario.

La clase *Email* encargada del envío de correos.

La clase *LogManager* consiste de un método que crea un *log header* y retorna un *string*

adecuado al proceso de *logging*.

La clase *Util* contiene métodos utilitarios que pueden ser requeridos a lo largo del código.

2.7.2.2. *Content Management API*: este paquete contiene clases Java para leer y manipular contenidos almacenados en el sistema de DSpace así como también los metadatos. Es la más usada por la capa de aplicación gracias a su correspondencia con los principales elementos del modelo de datos tales como: *Community*, *Collection*, *ítem*, *Bundle* y *Bitstream*. También contiene un paquete de *plugins* que permite la incorporación de nuevos ítems (contenidos) a DSpace así como la diseminación de esos contenidos.

2.7.2.3. *Plugin Manager*: es un contenedor simple de componente. Crea y organiza los componentes (*plugins*), y ayuda a seleccionarlos en los casos donde existan muchas opciones. Existen tres tipos de *plugins*: Singleton, Sequence y Named.

2.7.2.4. *Workflow System*: encargado de marcar los diferentes pasos para la incorporación de recursos. Más adelante se detallará este proceso.

2.7.2.5. *Administration Toolkit*: este paquete se encarga de algunas labores de administración tales como la creación de un usuario administrador, administración de registros de tipo *Dublin Core*, entre otros.

2.7.2.6. *E-person/Group Manager*: mantiene el registro de los diferentes usuarios a través de la creación y manipulación de las características que representan a los usuarios gracias a los métodos *get* y *set* para cada uno de esos atributos.

2.7.2.7. *Authorization*: como su mismo nombre lo dice, se encarga de gestionar las políticas de todas las autorizaciones para los ítems, comunidades, colecciones, grupos, usuarios, etc.

2.7.2.8. *Handle Manager/Handle Plugin*: este paquete tiene dos clases, una para crear y buscarlos los handle, y la otra clase para exponer y resolver los handle para sistemas externos a DSpace.

2.7.2.9. *Search*: representa una API simple la cual encapsula una máquina de búsqueda implementada con Apache Lucene.

2.7.2.10. *Browse API*: este paquete mantiene los índices de los atributos *dates*, autores, títulos y *subjects*.

2.7.2.11. *Checksum Checker*: es usado para verificar cada ítem dentro de DSpace.

2.7.2.12. *OpenSearch Support*: DSpace es capaz de soportar *OpenSearch* que es un pequeño conjunto de convenciones y documentos para describir y usar *search engines*, es decir, es una forma para que las páginas web y los motores de búsquedas publiquen los resultados en forma accesible.

2.7.2.13. *Embargo Support*: este paquete controla las labores de embargo de los recursos. Es importante destacar que el embargo es una restricción de acceso temporal al contenido.

2.7.3. Capa de Almacenamiento

2.7.3.1. *RDBMS*: DSpace usa una base de datos relacional que puede estar en PostgreSQL o en Oracle. El paquete `org.dspace.storage.rdbms` permite el acceso a través de sentencias SQL vía JDBC.

2.7.3.2. *Bitstream Store*: ofrece dos medios para almacenar contenidos, el primero es el sistema de archivo del servidor y el segundo es usando SRB (*Storage Resource Broker*).

2.8. Los Metadatos

Los metadatos son información estructurada que describe, explica y/o localiza un recurso de información para poder identificarlo, recuperarlo, describirlo, utilizarlo, administrarlo o preservarlo de una manera más clara y sistemática. Los metadatos son llamados en muchas ocasiones como datos acerca de los datos o información sobre la información. El término de metadatos se utilizan de manera diferente en distintas comunidades, algunas lo utilizan para referirse a la información de máquinas, mientras que otros lo hacen sólo para los registros que describen los recursos electrónicos. En el ambiente de una biblioteca, los metadatos se utiliza comúnmente para cualquier esquema formal de descripción de recursos de cualquier tipo digital o no digital. Estos esquemas de metadatos se han desarrollado para describir los distintos tipos de objetos textuales y no textuales, tales como libros, documentos electrónicos, herramientas de búsqueda de archivo, objetos artísticos, educativos y materiales didácticos o conjuntos de datos científicos.

Existen una gran variedad de esquemas de metadatos, uno de los primeros y más importantes para la consolidación de los metadatos en el mundo es el estándar internacional MARC (*Machine Readable Cataloging* o Catalogación Legible por Máquina) que se utiliza para la catalogación de libros y revistas de acuerdo con un esquema muy detallado que se define como conjunto de elementos de registro. MARC puede ser utilizado para la descripción de otros elementos distintos, pero presenta deficiencias en su capacidad para describir los requisitos de los objetos digitales, y ha sido suplantado en repositorios digitales por otras normas.

El *Dublin Core* ha resultado ser el más conocido y citado en los servicios de información en Internet por sus características [47], pues es simple y tiene un eficaz conjunto de elementos que describen una amplia gama de recursos, cada elemento es opcional y puede repetirse. La norma del *Dublin Core* conlleva dos niveles: Simple y Cualificado. El simple (DCS) presenta quince elementos y el cualificado presenta otros elementos adicionales al DCS. Sin embargo, con las múltiples dimensiones de los objetos digitales se están generando nuevos esquemas de metadatos adaptados a las diferentes dimensiones de los objetos tales como ID3, METS, MODS, MADS, TEI, MPEG-7, MPEG-21, LOM, entre otros.

De acuerdo con la Biblioteca de la Universidad de Cornell [48] y NISO Press [49] los metadatos se pueden clasificar en las siguientes categorías conforme a las funciones que cumplen:

1. *Descriptivos*: describen un recurso para poder describirlos y recuperarlos. Puede incluir elementos como el título, el resumen, autor y palabras clave.
2. *Estructurales*: refieren la estructura interna del recurso y los elementos que lo integran, facilitando la navegación y presentación de los recursos. Por ejemplo: página, sección, capítulo, numeración, índices, tablas de contenidos, entre otros.
3. *Administrativos*: proporcionan información para ayudar a gestionar un recurso, por ejemplo, datos técnicos sobre la creación, control de acceso y calidad, gestión de derechos, utilización y condiciones de preservación, migración, etcétera. Hay dos subconjuntos de datos administrativos, los derechos de gestión de metadatos y los metadatos de conservación.

2.9. Norma ISO 14721

En el año 2003 se publicó la Norma ISO:14721, denominada *ISO Reference Model for an Open Archival Information System (OAIS)* [50], que modela las partes que componen un sistema abierto de archivo de información, como por ejemplo un RI, sin definir su implementación. Las dos funciones principales del modelo OAIS son conservar la información y garantizar el acceso a la misma. La norma recomienda un entorno representado por un productor, un consumidor y un gestor:

- **Productor**: es quien provee la información para ser preservada y gestionada, pueden ser personas o sistemas de software.
- **Consumidor**: es quien busca, encuentra y adquiere información de interés que se ha preservado.
- **Gestor**: es el rol desempeñado por aquellos que establecen el conjunto de políticas del OAIS. No incluye las operaciones diarias.

El modelo OAIS propone un conjunto de seis módulos funcionales (también conocidos como entidades) que abarcan los diferentes servicios que pueden ser ofrecidos (figura 6). Para cada uno de estos módulos se definen unos requisitos funcionales así como la forma en que interactúan entre ellos. En la figura 6, se observan claramente los tres actores y las seis entidades.

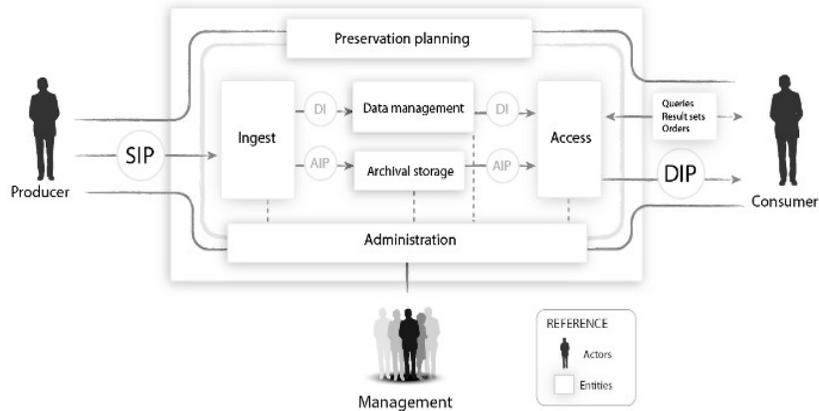


Figura 5. Modelo Funcional OAIS. Fuente: ISO 14721:2003

El elemento central de la norma es el *Information Package* (IP) lo conforman el objeto digital y todos sus metadatos, lo cual le permite moverse a través del modelo. El IP puede clasificarse en tres subtipos según su función en el proceso de archivo:

- *Archival Information Package* (AIP): contiene, como mínimo, suficiente información de un objeto como para garantizar la preservación a largo plazo. Busca mantener la mayor calidad posible de información descriptiva de preservación y de los objetos representados o contenidos.
- *Submission Information Package* (SIP): es el paquete que proviene del productor y se va a incorporar al OAIS. Suele contener menos información que el AIP.
- *Dissemination Information Package* (DIP): es el paquete que se entrega a un consumidor en respuesta a una solicitud. La información de empaquetado toma muchas formas dado que los usos de OAIS son diversos, puede ser tan completo como los AIP a partir de los cuales se construye o ser sólo una breve descripción del paquete.

El proceso (ver Fig. 6) puede iniciarse cuando el productor (un actor) suministra el recurso (paquete de entrada) llamado SIP a través del *ingest*, que luego se convierte en AIP terminando en la entidad de *archival storage*. El flujo puede continuar cuando el consumidor busca una información en el sistema, que es entregada como un DIP a través de la entidad *access*, ya que ha sido preservada en el sistema previamente. De forma simultánea los datos relacionados con los artículos y el repositorio mismo se mantienen organizados a través de la entidad de *data management* (gestión de datos). Luego, hay una entidad especializada para la administración adjunta a la gestión, que serían los administradores y el responsable del repositorio, y esta sección se relaciona con las secciones de *gestión de datos* y *planificación de la conservación*. Esto permite una gestión estructural y también ayuda a mantener los AIP a lo largo del tiempo. El módulo de *planificación de la conservación* desarrolla estrategias y normas de conservación, monitorea las últimas novedades y avances en el campo, además de monitorear los cambios en la comunidad designada, para que toda la información nueva que se solicite, se pueda adjuntar a los AIP correspondientes.

2.10. Preservación Digital

La preservación digital se basa en el conjunto de actividades necesarias para garantizar el acceso continuo a los recursos digitales durante todo el tiempo que sea necesario, sea cual sea el formato, software, hardware o sistema empleado para su creación. Combina las políticas, estrategias y acciones para garantizar el acceso independientemente de la tecnología. El objetivo de la preservación digital es la representación exacta del contenido en el tiempo.

La Guía Unesco para la preservación del patrimonio digital [51] indica que “...el patrimonio digital está constituido únicamente por aquellos que se considera que poseen un valor permanente”. Actualmente existen numerosas estrategias para asegurar la preservación de la información y evitar los problemas de obsolescencia, entre los que podemos mencionar: migración continúa, adhesión a estándares internacionales, emulación, políticas de *backups*, metadatos de preservación, encapsulamiento...

Tránsito Ferreras Fernández [52] afirma que los esfuerzos para preservar los objetos digitales no se acaban con lo técnico, sino que existen otros: legales (permiso de los autores), económicos e institucionales (compromiso permanente) para llegar al resultado de perdurabilidad.

Catherine Lupovici y Julien Masanés [53] definen un conjunto mínimo de metadatos de carácter obligatorio para el propósito de preservación en un contexto de grandes volúmenes y tecnología variable. Proponen metadatos y sub-elementos siguiendo la taxonomía del modelo OAIS. Proponen que la generación de metadatos se realice, en lo posible, de manera automática.

El proyecto denominado *Repository preservation and interoperability* (Preserv2) [54], se basa en el servicio de PRONOM provisto por *The National Archives* (TNA) y la herramienta DROID (*Digital record object identification service*) [55] que usa los perfiles de formato de más de 200 repositorios del registro PRONOM.

JISC continuó el proyecto PReserv2 con el proyecto Keepit [56] para permitir gestión eficaz y mantenible a lo largo del tiempo de una amplia gama de contenido digital presente en los repositorios institucionales: artículos de investigación, datos primarios, artes, materiales didácticos y tesis.

2.10.1. Metadatos de Preservación

Los metadatos de preservación soportan los datos necesarios para cumplir con una serie de requerimientos de preservación con el objetivo de asegurar la utilización a largo plazo de un recurso digital. Algunos de los requerimientos son:

1. Debe mantenerse el objeto digital en el repositorio de manera segura sin perderse ni ser modificado sin autorización.
2. Se debe conocer su creador.

3. Si cambia se debe conocer quién realizó el cambio.
4. Debe poder localizarse y entregarse al usuario.
5. Debe almacenarse en soportes que puedan leer los sistemas actuales de manera que el usuario pueda comprenderlos.
6. Del mismo modo las estrategias de emulación y migración requieren metadatos sobre los formatos de los objetos originales y los entornos de hardware y software que los soportan.
7. Soportar la autenticidad mediante la documentación de la *procedencia digital* a través de su cadena de custodia y el historial de cambios autorizados.
8. El repositorio debe disponer de los derechos suficientes como para llevar adelante las transformaciones necesarias para mantener el acceso al objeto.
9. Si el objeto está relacionado con otros del repositorio o de otros depósitos externos, estas relaciones deben guardarse.

En resumen, los metadatos de preservación están destinados a almacenar los detalles técnicos sobre el formato, la estructura, el acceso y el uso de los contenidos digitales, la historia de todas las acciones realizadas en el recurso, incluyendo los cambios, la información de autenticidad, las características técnicas o la historia de la custodia y las responsabilidades y la información sobre los derechos con que se cuenta para realizar las acciones de preservación. Muchos repositorios representan la información de preservación bajo PREMIS, METS o una combinación de ambos. En las siguientes secciones se realiza una breve descripción de cada uno.

2.10.2. PREMIS

PREMIS [57] es un grupo internacional patrocinado por *Online Computer Library Center (OCLC)* y *Research Libraries Group (RLG)* que, como su nombre lo indica, se enfoca en estrategias de implementación de metadatos y preservación en Archivos Digitales. En 2008, este grupo elaboró el Diccionario de Datos PREMIS para Metadatos de Preservación [58], [59], que define los metadatos de preservación como “la información que utiliza un repositorio para dar soporte al proceso de preservación digital”. El diccionario define un conjunto de *unidades semánticas*, propiedades e información que la mayoría de los repositorios necesita conocer de sus entidades para asegurar la preservación.

Los metadatos PREMIS se concentran sólo sobre los elementos que afectan directamente a la preservación, sobre el resto de los metadatos no se mantiene vinculación ni al acceso, ni a la recuperación de la información, ni siquiera a la información sobre derechos. PREMIS plantea la necesidad de representar las unidades semánticas de forma abstracta, aunque no regula su implementación ni representación. Las unidades semánticas de PREMIS se corresponden con los metadatos definidos en el esquema PREMIS XML y pueden corresponderse parcialmente con algunos de los formatos más usados.

2.10.3. METS

METS [60] está desarrollado por *Network Development and MARC Standards Office* de la *Library of Congress*. Es un formato que registra la estructura jerárquica de un objeto digital: nombre, archivos, ubicación, estructura y metadatos asociados. Un documento

METS posee un formato estandarizado para transmisión de metadatos o archivos de datos, que se estructura XML. Eva Méndez lo define como una “estructura contenedora” extensible y modular que utiliza extensiones “wrappers” o “sockets” donde los elementos de otros esquemas pueden conectarse [61]. METS utiliza el esquema XML que le resulta útil para combinar distintos espacios de nombre. Un documento METS consta de siete secciones:

1. Cabecera METS - contiene metadatos que describen el propio documento: identificación, creador, fecha de creación y agentes.
2. Metadatos Descriptivos: Esta sección puede: a) enlazar a metadatos descriptivos externos al documento METS; b) contener internamente los metadatos descriptivos posibles de registrar en formato XML o cualquier otro formato binario o textual, o c) combinar ambas modalidades.
3. Metadatos Administrativos: contiene metadatos técnicos con información relativa a la creación de los archivos que conforman al objeto digital, su formato y características de uso, sobre *copyright* e información sobre licencias, sobre el origen, y sobre la procedencia digital, es decir, sobre la relación entre el documento original y su representación digital, incluyendo la relación entre copias maestras y derivadas, migraciones y transformaciones.
4. Archivo: lista todos los archivos con contenidos que forman parte del objeto digital.
5. Mapa Estructural: recoge la estructura jerárquica del objeto digital, enlaza sus secciones con los archivos de contenido y los metadatos correspondientes a cada una de ellas.
6. Enlaces Estructurales: permite registrar la existencia de hiperenlaces entre las secciones del mapa estructural.
7. Comportamientos: permite asociar el contenido del objeto con *comportamiento ejecutable*. Cada comportamiento posee un elemento que define su *interface* y un *mecanismo*, que identifica un módulo de código ejecutable que implementa y ejecuta los comportamientos definidos por la *interface*.

2.10.4. METS-PREMIS

El par METS-PREMIS es muy adecuado a los fines de la preservación, los esquemas más utilizados de metadatos cuentan con metadatos útiles para ese cometido. Es importante tratar de detectar en el esquema usado por el repositorio, aquellos que resulten útiles para proveer datos técnicos sobre el formato, la estructura y el uso de los contenidos digitales, la historia de todas las acciones realizadas en los recursos, los cambios, la autenticidad, las responsabilidades y los permisos.

En algunos esquemas pueden surgir ambigüedades (algunas ampliamente referenciadas), sin embargo, una práctica más elaborada a la hora de incorporar los metadatos (por ejemplo, el agregado de subelementos en el esquema DC) puede ayudar a la desambiguación. Es claro que los esquemas que no están dedicados, de manera específica, a la preservación tienen limitaciones, aunque pueden ser un buen comienzo para contar con datos de preservación e incorporar criterios para adoptar mejores prácticas. A continuación se listan algunos metadatos de DC Simplificado [47] y MODS [62] que resultan útiles en la preservación según PREMIS [57], ver tablas 1 y 2:

DUBLIN CORE	Elemento PREMIS	DEFINICIÓN
Identifier	Object Entity (1.1 objectIdentifier)	Referencia única al recurso en un contexto determinado
Creator	Agent Entity	Responsable de crear el recurso
Description	Object Entity (1.10 relationship)	Resumen del recurso
Date	Object Entity (1.5.5 creatingApplication)	Tiempo asociada con un evento en el ciclo de vida del recurso
Format	Object Entity (1.5.4 format)	Presentación física o digital del recurso
Relation	Object Entity (1.10 relationship)	Una referencia a un recurso asociado
Rights	Right Entity (4.1 RightsStatements)	Información sobre los derechos legales que afectan el uso del recurso

Tabla 2. DC-PREMIS

MODS CORE Element	Elemento PREMIS	DEFINICIÓN
Identifier	Object Entity (1.1 objectIdentifier)	Número único que identifica un recurso
name	Agent Entity	Asociado de alguna forma con el recurso
internetMediaType	Object Entity (1.5.4 format)	Presentación física o digital del recurso
relatedItem	Object Entity (1.10 relationship)	Resumen del contenido del recurso
location	Object Entity (1.7.1 contenLocation)	Identifica la institución o repositorio que tiene el recurso o ubicación remota a través de un URL
accessCondition	Right Entity (4.1 RightsStatements)	Información sobre las restricciones impuestas al acceso a un recurso

Tabla 3. MODS-PREMIS

2.11. Estándares a Tomar en Cuenta

2.11.1. XML

XML es un lenguaje de marcas desarrollado por el *World Wide Web Consortium (W3C)*, conocido en inglés como *eXtensible Markup Language*. Deriva del lenguaje SGML (ISO 8879) y permite definir la gramática de lenguajes específicos (de la misma manera que HTML es a su vez un lenguaje definido por SGML) para estructurar documentos electrónicos de gran escala; juega un papel importante en el intercambio de una amplia variedad de datos en la red dispuestos en cualquier lugar. XML se creó para que fuera idéntico a la forma de recibir y procesar la información del HTML, extensible y fácil de implantar, programar y aplicar a los distintos sistemas.

XML es un lenguaje que permite describir los contenidos dentro del propio documento, reutilizar sus partes, jerarquizar y estructurar esos contenidos. La información estructurada presenta varios contenidos (texto, imágenes, audio, etc.) y formas (hojas de cálculo, tablas de datos, parámetros de configuración, etc). La forma indica qué papel puede jugar el contenido, por ejemplo, el contenido de una sección encabezada con un significado difiere del contenido de una nota a pie de página, o el contenido de un pie de foto difiere del contenido de una tabla de datos. Otro ejemplo es la codificación en *Dublin Core* en un esquema XML que permite que los motores de búsqueda puedan encontrar información de forma mucho más eficiente, permitiendo el intercambio de información entre diversos sistemas diferentes por presentar un formato estándar para representar la información sobre un tema específico.

2.11.2. OAI-PMH

El protocolo OAI-PMH (*Open Archives Initiative-Protocol Metadata Harvesting*) realiza el intercambio de información para que desde diversos puntos (proveedores de servicio) se puedan hacer búsquedas que abarquen la información recopilada en distintos repositorios asociados (proveedores de datos), tal y como se observa en la figura 7.

El volumen y crecimiento de archivos y de motores de búsqueda han hecho que se recurra a los metadatos y al lenguaje estructurado de marcas XML para conseguir la interoperatividad entre archivos. Esta iniciativa permite el *harvesting* de los metadatos, que se entiende como la localización y recopilación de los metadatos que describen los registros contenidos en los repositorios mediante el protocolo OAI-PMH. Los metadatos a intercambiar deben codificarse en *Dublin Core Simple* (no calificado).

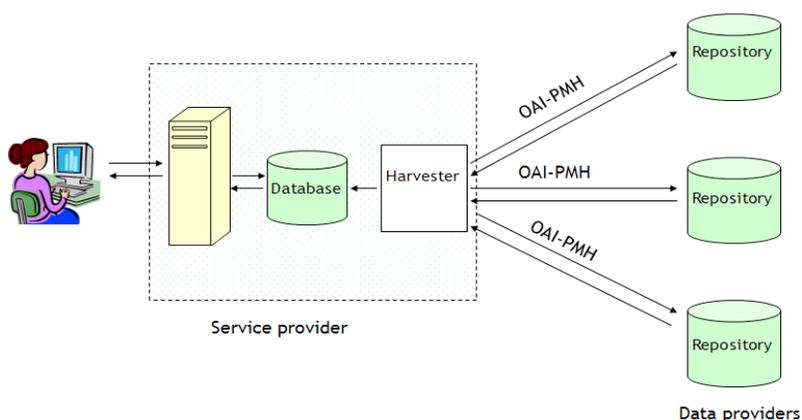


Figura 6. Proveedores del Protocolo OAI-PMH

2.11.2.1. *Proveedores de datos*: son los depósitos de documentos que proporcionan los metadatos de los recursos que almacenan. Un proveedor de datos ofrece los metadatos a través de un archivo XML siguiendo una estructura específica de los recursos resguardados (acceso público o privado).

2.11.2.1. *Proveedores de servicios*: son conocidos como recolectores o agregadores de

contenidos y se definen como un programa encargado de recolectar la información ofrecida por los proveedores de datos con el objetivo de incorporarles algún valor añadido y presentarlos a los usuarios finales.

2.11.3. SWORD

SWORD (*Simple Web-service Offering Repository Deposit*) es un estándar de interoperabilidad que permite a los repositorios digitales aceptar el depósito de los contenidos de múltiples fuentes en diferentes formatos (por ejemplo, documentos XML) a través de un protocolo estandarizado. De la misma manera que el protocolo HTTP permite que cualquier navegador web para hablar con cualquier servidor web, de modo SWORD permite a los clientes hablar con los servidores de repositorio.

2.11.4. OpenSearch

OpenSearch es un conjunto de tecnologías que logra la publicación de los resultados de búsqueda en un formato adecuado para la sindicación y agregación. Es una manera para que los sitios web y motores de búsqueda publiquen los resultados de búsqueda en un formato estándar y accesible. *OpenSearch* 1.0 fue publicado en marzo de 2005. Consiste en cinco pilares: archivos descriptivos, sintaxis de consulta, RSS, agregadores y un "Auto-discovery".

2.11.5. RSS

RSS (*Really Simple Syndication*) es un formato XML para indicar o compartir contenido en formatos de fuentes web utilizadas para publicar frecuentemente actualización de productos de software, tales como entradas de *blogs*, titulares de noticias, audio y video en un formato estandarizado. Un documento RSS incluye un resumen o texto completo, además de los metadatos como fechas de publicación y autoría.

Existen editores y lectores de RSS. Los editores distribuyen el contenido de forma automática y los lectores, a través de suscripciones vía URI, permiten a los usuarios evitar manualmente inspeccionar todos los sitios web que están interesados, de tal manera que todo el contenido nuevo se inserta en su programa lector (principalmente el navegador) cuando esté disponible.

2.11.6. Search/Retrieve Web service (SRW)

Es un servicio web para la búsqueda y recuperación de información que proporciona una interfaz SOAP para las consultas. Este estándar es creado por la Biblioteca del Congreso de los Estados Unidos.

2.11.7. Search/Retrieve vía URL (SRU)

Es un protocolo estándar de búsqueda para las consultas de búsqueda de Internet, utilizando lenguaje de consulta contextual (CQL), una sintaxis de consulta estándar para la representación de las consultas. Este estándar y el CQL son creados por la Biblioteca del Congreso de los Estados Unidos.

2.12. Marco Legal en la República Bolivariana de Venezuela

La propuesta presentada se acoge al marco legal presente en la República Bolivariana de Venezuela de acuerdo con la Constitución de la República Bolivariana de Venezuela de 1999 y con las leyes, reglamentos y decretos promulgados en el país. A continuación, se enumeran los artículos en los que se respalda la propuesta en cada uno de los documentos legales:

- Constitución de la República Bolivariana de Venezuela de 1999, artículo 110: *"El Estado reconocerá el interés público de la ciencia, la tecnología, el conocimiento, la innovación y sus aplicaciones y los servicios de información necesarios por ser instrumentos fundamentales para el desarrollo económico, social y político del país, así como para la seguridad y soberanía nacional. Para el fomento y desarrollo de esas actividades, el Estado destinará recursos suficientes y creará el sistema nacional de ciencia y tecnología de acuerdo con la ley. El sector privado deberá aportar recursos para las mismos. El Estado garantizará el cumplimiento de los principios éticos y legales que deben regir las actividades de investigación científica, humanística y tecnológica. La ley determinará los modos y medios para dar cumplimiento a esta garantía"*.
- Ley Orgánica de la Administración Pública promulgada en el 2001. El artículo 12 insta, de forma explícita, al uso de las nuevas tecnologías en la administración pública: *"...A fin de dar cumplimiento a los principios establecidos en esta Ley, los órganos y entes de la Administración Pública deberán utilizar las nuevas tecnologías que desarrolle la ciencia, tales como los medios electrónicos, informáticos y telemáticos, para su organización, funcionamiento y relación con las personas..."*.
- Ley Orgánica de Educación promulgada en el 2009, el artículo 27 expresa que la educación debe ser de libre acceso: *"La educación intercultural transversaliza al Sistema Educativo y crea condiciones para su libre acceso a través de programas basados en los principios y fundamentos de las culturas originarias de los pueblos y de comunidades indígenas y afrodescendientes..."*.
- Ley Orgánica del Trabajo de los Trabajadores y Trabajadoras promulgada en el 2012. El artículo 325 establece: *"La producción intelectual generada bajo relación de trabajo en el sector público o financiada a través de fondos públicos que origine derechos de propiedad intelectual, se considerarán del dominio público, manteniéndose los derechos al reconocimiento público del autor o autora"*.
- Ley y Reglamento sobre Mensajes de Datos y Firmas Electrónicas promulgados en el 2001 y en el 2004, respectivamente.
- Ley sobre Acceso e Intercambio Electrónico de Datos, Información y Documentos entre los órganos y Entes del Estado promulgada en el 2012.
- Reglamento de la Ley sobre Mensajes de Datos y Firmas Electrónicas del 2004.
- Decreto 825 promulgado en el 2000 que sobre la importancia de Internet.
- Decreto 3390 promulgado en el 2004, que indica el uso prioritario de Software Libre de Estándares Abiertos en la Administración Pública Venezolana en sus sistemas, proyectos y servicios informáticos.

2.13. Control de Autoridades de Registros Bibliográficos

Una autoridad es una forma normalizada de cualquier entrada de un catálogo bibliográfico, registrada en un fichero manual o automatizado. Este registro suele contar con datos sobre los términos no seleccionados y notas de aplicación y de fuente. Según el Glosario ALA [63], un archivo de autoridad es un "Conjunto de registros normalizados que establece las formas autorizadas de los encabezamientos que han de utilizarse en un conjunto de registros bibliográficos y las referencias que deben hacerse a los encabezamientos de éstos". El registro de autoridades es el "Registro en el que aparece un encabezamiento establecido para su empleo en un conjunto de registros bibliográficos, que cita las fuentes consultadas al establecer el encabezamiento, indica las referencias que han de hacerse a y desde éste, expresando la información encontrada en las fuentes como justificación de la forma de encabezamiento elegido y de las referencias especificadas". Algunos autores distinguen entre trabajo de autoridades y control de autoridades, considerando el primero más amplio, pues engloba la investigación previa y la creación de nuevas autoridades y sus relaciones; mientras que el control de autoridades es el mantenimiento de la consistencia de los encabezamientos de un fichero bibliográfico mediante las referencias del archivo de autoridades.

2.13.1. Clasificación

Existen tres grupos generales de autoridades de acuerdo con su creación y las áreas del registro bibliográfico en que se consignan [62]:

- Las **autoridades de nombre** son las que corresponden a las menciones de responsabilidad de la obra, por tanto, al generarlas se tiene muy en cuenta la forma en la que aparece en la publicación y se caracterizan. Por ejemplo: autor, título, autor corporativo, congreso, etc.
- Las **autoridades de materia** se caracterizan porque habitualmente las formas anteriores de un término se trazan como entradas de referencia o términos alternativos; es decir, no suelen coexistir dos formas de un mismo término. Por ejemplo: la materia propiamente, lugar geográfico, género, forma, etc.
- Las **autoridades de serie** son las peor conocidas y las de más complejidad, debido a las propias características de éstas y a la escasez de normalización. Pueden ser autoridades de serie un nombre de persona con un título, una entidad con un título o un título uniforme. Afectan al área de serie y su objetivo es agrupar los distintos títulos publicados en una serie o colección.

Capítulo 3

Metodología

Para el desarrollo del presente se realizó un tipo de investigación de la modalidad de Proyectos Especiales según el Manual de la UPEL [64], porque la creación tangible y apropiada podrá ser utilizada por una institución para la gestión y preservación de sus documentos digitales administrativos. Este trabajo posee un carácter innovador y representa un aporte al área de los Repositorios Digitales. Se estructuró en tres fases que serán detalladas a continuación.

3.1. Primera Fase: Relevamiento Bibliográfico

En la primera fase se realizó una contextualización del problema planteado en la UNET, los objetivos de la propuesta, la motivación, los alcances y limitaciones, y una revisión bibliográfica de los siguientes aspectos:

- Bibliotecas Digitales.
- Repositorio Institucional.
- Repositorios de Documentos Administrativos.
- Metadatos.
- Acceso Abierto.
- Preservación Digital.
- Marco Legal para los Repositorios Institucionales en Venezuela.
- Control de autoridades.

3.2. Segunda Fase: Análisis e Implementación

Esta fase comprende el análisis del problema e implementación del REDAUNET, con el fin de obtener una solución al problema planteado. Los elementos que se tomaron en cuenta fueron:

- Planificación
- Instalación
- Configuración
- Plataforma de trabajo
- Catalogación

3.3. Tercera Fase: Documentación Final de la Propuesta

Tercera Fase: Documentación Final de la Propuesta

En esta fase se construyó la Wiki SEDICI-DSpace, que profundizó de forma técnica cada uno de los elementos de la propuesta. La estructura de la Wiki se puede consultar en el Apéndice B. En esta fase se desarrollaron las consideraciones finales como recomendaciones, conclusiones y trabajos futuros .

Capítulo 4

Implementación

4.1. Planificación

En las secciones anteriores se ha observado cómo han evolucionado los sistemas para los Repositorios Institucionales. Por ello, se realizaron búsquedas de las herramientas más usadas según los directorios de repositorios existentes, como OpenDOAR y ROAR, entre los cuales se destacan DSpace, FEDORA y EPrints. Luego, se hizo un análisis de los siguientes puntos: soporte y mantenimiento del software, interfaces gráficas y configurables, flexibilidad en los sistemas, configuración de los formatos de metadatos, soportes para interoperabilidad, sistemas para la preservación y *backup*, procedimientos de instalación, entre otros. Gracias al estudio que realizó el equipo de trabajo de la Universidad Nacional de La Plata mediante SeDiCI (Servicio de Difusión de la Creación Intelectual) se decidió usar DSpace.

SeDiCI, desde sus inicios (2003), tomó las políticas del software libre y el *Open Access*. Aporta desarrollos e investigaciones sobre esos lineamientos y su misión es albergar, preservar, difundir y dar visibilidad a nivel mundial de la producción científica e intelectual de las distintas unidades académicas que la componen. Por ello, a partir de la experiencia de SeDiCI y de acuerdo con el manual de LEADIRS II de Barton y Waters [65], se pueden tomar en cuenta los siguientes pasos para la implementación de repositorios:

- Poner en contexto el concepto de Repositorios.
- Desarrollar una definición y un plan de servicio:
 - Realizar una evaluación de las necesidades de su Universidad.
 - Desarrollar un modelo de coste basado en este plan.
 - Crear una planificación y un horario.
 - Desarrollar políticas de actuación que gestionen la recopilación de contenidos, su distribución y mantenimiento.
- Conformar el equipo de trabajo.
- Elegir la tecnología adecuada.
- Difundir el servicio.
- Poner en funcionamiento el repositorio.

Estas pautas son básicas y pueden variar de acuerdo con cada institución, pero ayudan a marcar una línea lógica para la planificación. Esta propuesta se limitará a indicar los pasos para una instalación, configuración básica, un entorno de trabajo mínimo, opciones de catalogación y la documentación recomendada.

4.2. Instalación

4.2.1. Entorno de Desarrollo

El ambiente de desarrollo propuesto surgió del Manual de DSpace 1.8. que expresa la necesidad de independizar la creación de componentes propios sin modificar el *Core* de

DSPACE, para garantizar una continuidad de versiones del software. Por tal motivo, se debe contar con un sistema de control de versiones. A continuación se muestra la configuración mínima sugerida para tener un entorno de desarrollo y un mantenimiento adecuado del repositorio que se desea:

1. Sistema operativo: Linux Debian.
2. Java Development Kit (JDK) 1.6 o superior. Evitar OpenJDK.
3. Entorno de Desarrollo Integrado (IDE) Eclipse Indigo 3.7 o Eclipse Juno 4.2.
4. Tomcat 6.x o superior.
5. PostgreSQL.
6. Apache Maven 2.2.x.
7. Requisitos de Hardware: arquitectura Intel. Procesador Core i3 (o equivalente) como mínimo. Memoria RAM mínima de 2 GB. Disco duro con 5 GB libres como mínimo.
8. Una herramienta para la gestión de proyectos de software y seguimiento de errores, por ejemplo Redmine o TRAC.
9. Plugins para Eclipse:
 - SVN (subversive o subclipse).
 - m2e/Maven Integration for Eclipse.
 - Maven SCM Handler for Subversive, agregar el repositorio que corresponda según el plugin instalado de SVN.
 - Instalar el plugin Sysdeo Tomcat.

4.2.2. Recomendación *One Big Project* de DSPACE

La documentación y la estructura de POMs del proyecto DSPACE establecen dos posibles puntos de extensión: *DSPACE Modules* y *Application Overlays*, gracias al Maven y Ant. El Maven es necesario para construir y ensamblar el paquete DSPACE que incluye la aplicación de *overlays* y dependencia de terceros, a través de los `pom.xml`. El Ant es necesario para la instalación y actualización del directorio de instalación del DSPACE, `[dSPACE.dir]/config/dSPACE.cfg`. *DSPACE Modules* se evidencia a través de un nuevo módulo (otra aplicación web, una aplicación en *swing*, una librería, etc.) creando un directorio en la raíz del proyecto, el cual debe incluir un archivo `pom.xml` y luego terminar de construir la estructura de directorios que corresponda. La extensión *Application Overlays* está basada en procesos *overlays*, los cuales se encargan de unir los desarrollos propios con los desarrollos de DSPACE, que por lo general se localizan en el directorio `/dSPACE/modules/`, donde cada módulo genera una aplicación web (archivo `.WAR`) con los elementos de personalización que se consideren necesarios en cada caso.

Después de definir la forma de hacer la extensión de los componentes deseados, DSPACE recomienda trabajar directamente en Eclipse como un proyecto SVN o un proyecto SCM de Eclipse, convirtiéndose en un único proyecto padre que contiene todos los módulos de DSPACE y los desarrollados propiamente. La estructura del proyecto desde el SVN de DSPACE, `svn.duraspace.org/dSPACE/dSPACE/trunk`, es:

```
/dSPACE/ ==> DSPACE 'build' and configuration module
```

```

.... /bin/
.... /config/
.... /docs/
.... /etc/
.... /modules/
..... /jspui/
..... /lni/
..... /oai/
..... /solr/
..... /sword/
..... /sword2/
..... /xmlui/
..... /pom.xml
.... /solr/
.... /src/
.... /pom.xml
/dspace-api/ ==> Java API source module
/dspace-discovery/ ==> Discovery source module
/dspace-jspui/ ==> JSP-UI source module
/dspace-lni/ ==> Lightweight Network Interface source module
/dspace-oai/ ==> OAI-PMH source module
/dspace-stats/ ==> Statistics source module
/dspace-sword/ ==> SWORD deposit service source module
/dspace-sword-client/ ==> XMLUI client for SWORD
/dspace-swordv2/ ==> SWORDv2 source module
/dspace-xmlui/ ==> XML-UI (Manakin) source module
/pom.xml ==> DSpace Parent Project definition

```

4.2.3. Ant - Maven

Una vez instalado el software recomendado y creado un proyecto SVN o SCM en Eclipse, se realizarán una serie de tareas repetitivas a nivel de consola tales como: borrar todos los archivos de extensión `class` para recompilar desde cero, compilar, generar la documentación de nuestro programa con `javadoc`, generar el `jar` con nuestro proyecto, etc. Estos procesos pueden hacerse con Windows a través de archivos `.bat` o en Linux con `scripts`, pero ninguno de ellos presenta portabilidad para diversos sistemas operativos, por ello, se creó Ant y Maven.

Ant es una herramienta gratuita que funciona similar a los *scripts* o archivos `.bat`. En un archivo XML se ponen las tareas que se deseen ejecutar, el archivo normalmente se llama `build.xml`. Luego, desde la línea de comandos, ya sea en Windows o Linux se ejecuta el comando `ant` y se inicia el proceso, ya sea la compilación, borrado o generación de documentación. En Ant se realiza en Java, permitiendo la portabilidad en los diferentes sistemas operativos donde se desarrolle el repositorio.

En cuanto a Maven, es una herramienta más evolucionada, ya que crea una estructura de directorios para el proyecto con sitio para los fuentes, los iconos, ficheros de configuración y datos, etc. Maven puede recibir las instrucciones de las librerías (.jar) externas que se necesitan, y es capaz de ir a buscarlos a Internet para descargarlos por nosotros. También, sin necesidad prácticamente de configurar nada, Maven sabe cómo borrar los .class, compilar, generar el jar, buscar las dependencias, distribuir y generar la documentación. El archivo que trabaja normalmente con Maven es pom.xml. En el proyecto de DSpace, en Eclipse, se pueden encontrar los archivos build.xml y pom.xml en el directorio raíz del proyecto,/[dspace.dir]/.

4.2.4. Recomendación de Proyecto SeDiCI-DSpace

La experiencia del personal de SeDiCI condujo a una recomendación para el desarrollo y mantenimiento del repositorio deseado. Por ello, a partir de la recomendación de *One Big Project* de DSpace se ofrece un SVN de SeDiCI, svn.prebi.unlp.edu.ar/SeDiCI-DSpace/trunk, resultando la siguiente estructura:

```
/distribution/  
.../dspace-tag-files/  
.../overwritten-files/  
.../pom.xml  
/install/  
/plugins/  
/sedici2003/  
/sedici2003-api/  
/xmlui/  
.../src/main/java/  
.../src/main/resources/  
.../src/main/webapp/  
.../pom.xml  
/build.bash  
/build.defaults  
/config.prop  
/import.bash  
/pom.xml
```

La estructura está diseñada para trabajar vía extensión *Overlay*, separando en proyectos propios los cambios y componentes nuevos desarrollos. Luego al compilar, los desarrollos propios se unen al proyecto DSpace vía *Overlay*. Por tanto, los subproyectos mostrados anteriormente son:

- /dspace-tag-files/: una copia del SVN de DSpace.
- /overwritten-files/: clases o archivos de configuración que han sido modificados.
- /install/: directorio que contiene la instalación del DSpace-SeDiCI y que luego es levantado como aplicación para el usuario. Este proyecto es el que se debe

montar en el servidor de PRODUCCIÓN.

- /plugins/: desarrollos propios.
- /sedici2003/ y /sedici2003-api/: desarrollo que enlaza con el sistema de repositorios previo al DSpace-SeDiCI actual
- /xmlui/: interfaz de usuario con el que trabaja nuestro repositorio.
- El archivo `build.bash` es el script encargado de compilar el proyecto DSpace-SeDiCI.

4.2.5. Interfaz Gráfica de Usuario: XMLUI

Conocida como Manakin, está desarrollada a partir del *framework* Cocoon y tiene tres componentes:

1. Aspectos.
2. Temas.
3. Esquema DRI.

4.2.5.1. *Aspectos*: es la característica o la funcionalidad que se agrega al *front* de forma independientes bajo la filosofía del SoC (*Separation of Concerns*), p.e. *artifact browser*, *e-persons*, *submission*, etc.

4.2.5.2. *Temas*: representan la forma de presentación del *front* y se pueden ubicar en el `path` `/[source.dspace.dir]/themes/`, por ejemplo: Classic, Kubrick, Mirage, Sedici, etc.

4.2.5.3. *Esquema DRI*: es una representación abstracta que propone DSpace para generar la página web a mostrar al usuario. El documento DRI contiene información (metadatos) que será mostrada en la página web (como un documento XHTML), ver la siguiente figura:

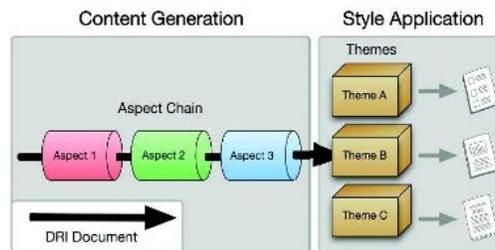


Figura 7. Esquema DRI. Fuente: DSpace

Cocoon está diseñada alrededor del concepto de separación de preocupaciones (SoC), que usa una arquitectura basada en componentes. Se concentra en la publicación de *websites* utilizando SAX (Simple API for XML) y poniendo fuerte concentración en los *pipeline*.

El principio de SoC es realizado a través de componentes que unidos forman los *pipelines*. Algunos de los componentes más utilizados son: *matchers*, *generators*, *transformers*, *serializer*, *selectors*, *views*, *readers* y *actions*. Dichos componentes no interactúan entre sí, sin embargo, su interacción es guiada por su constructor, el cual, en la mayoría de los casos, es representado con el archivo `sitemap.xml`.

El código fuente de esta interfaz se encuentra en la ruta `/[dspace.dir]/xmlui/` y la estructura del paquete sería:

```
/src/  
  ....main/  
  .....java/  
  .....resources/  
  .....webapp/  
  ....test/  
/target/  
  ....classes/  
  ....generated-sources/  
  ....maven-archiver/  
  ....war/  
  ....xmlui-1.8.2/  
  ....xmlui-1.8.2.war  
/pom.xml
```

4.2.6. Verificación de la Instalación

El administrador necesita verificar la instalación asegurando que todos los componentes estén funcionando. A continuación se lista la verificación mínima que se debe realizar:

- El sistema debe estar funcionando correctamente, entonces, debe estar funcionando el servidor web (en la propuesta es Apache Tomcat) y el sistema debe estar accesible.
- La base de datos debe estar corriendo y funcionando correctamente, y debe poderse crear un usuario, una comunidad o una colección. Se recomienda ejecutar el comando de prueba para base de datos de DSpace de la siguiente manera: `/[dspace]/bin/dspace test-database.`
- Sistema de correos electrónico funcionando. En caso de alguna falla, revisar la configuración en `dspace.cfg` y hacer el test de correos de DSpace: `/[dspace]/bin/test-email.`

4.3. Configuración

En esta sección se describe cómo configurar los módulos a tener en cuenta para la implementación del repositorio:

1. XMLUI
2. Metadatos, proceso de submission y flujos de trabajo (workflow)
3. Comunidades y colecciones
4. Control de autoridades
5. Tipos de documentos
6. OAI-PMH
7. Discovery
8. Usuarios y Grupos

Sin embargo, antes de describir cada uno de ellos, es importante destacar que el archivo principal de configuración de Dspace es `/[dspace-source]/dspace/config/dspace.cfg`. Este archivo está compuesto por las propiedades del sistema con sus valores por defecto. El conjunto de propiedades básicas más destacadas son:

- Basic Information
- Database Settings
- Advanced Database Configuration
- Email Settings
- File Storage
- Logging Configuration
- Search Settings
- Handle Settings
- Proxy Settings
- Embargo Settings
- Checksum Checker
- Submission Process
- Browse Configuration
- Syndication Feed (RSS) Settings
- Sitemap Settings
- Authority Control Settings
- XMLUI Settings (Manakin)
- SOLR Statistics Configurations

4.3.1. Personalización de XMLUI

El resultado de seleccionar la interfaz web XMLUI de DSpace se puede observar en el repositorio institucional de la Universidad Nacional de La Plata, para octubre del 2012, con un tema adaptado a este tipo de interfaz. La distribución de esta interfaz se presenta a continuación:

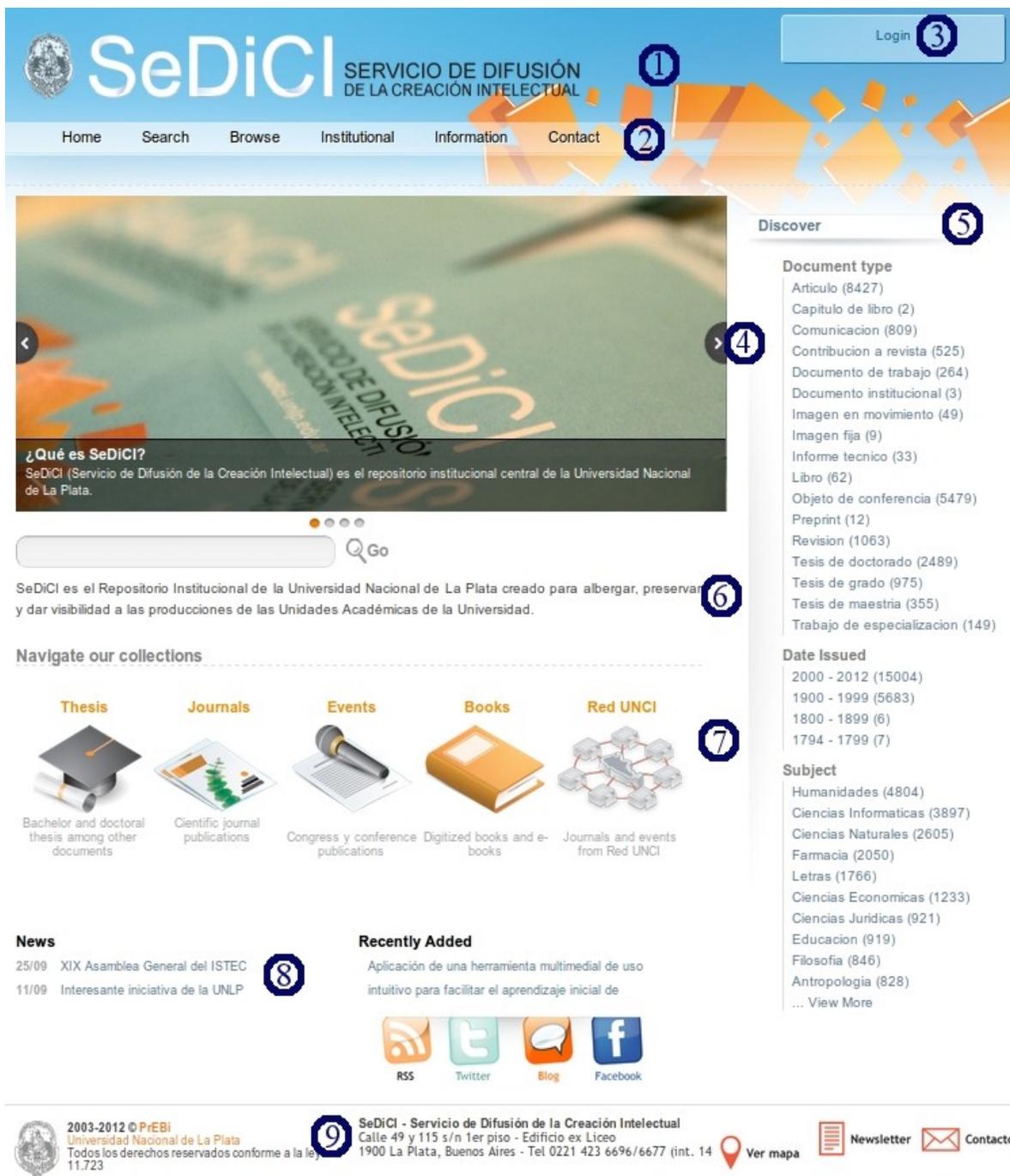


Figura 8. Interfaz principal de SeDiCI, octubre 2012

SeDiCI pone a disposición este tema XMLUI de DSpace, entonces, para lograr esta configuración se indican los archivos que corresponden a cada una de las partes numeradas en la figura anterior:

1. Parte superior del portal:
/xmlui/src/.../style.css

2. Menú superior:
 /xmlui/src/.../menu-superior.xml
 /install/config/pages/*.xml
3. Ingreso de usuarios registrados: es un aspecto que es configurado en el directorio correspondiente ubicado en `xmlui/src/main/resources/aspects/` y habilitándolos en el archivo de configuración `/distribution/overwritten-files/config/xmlui.xconf`.
4. Sliceshow:
 /xmlui/src/.../slideshow.xml
5. Discover: igual que el proceso de ingreso de usuarios, es un aspecto que es configurado en el directorio correspondiente ubicado en `xmlui/src/main/resources/aspects/` y habilitándolos en el archivo de configuración `/distribution/overwritten-files/config/xmlui.xconf`.
6. Reseña del portal:
 /install/config/news-xmlui.xml
7. Menú de navegación por colecciones:
 /xmlui/src/.../Home.xml
8. Noticias:
 /xmlui/src/.../news/news-list.xml
 /xmlui/src/.../ShowNews.java
 /install/config/modules/sedici-dspace.cfg
9. Parte inferior del portal:
 /xmlui/src/.../page-structure.xml
10. Finalmente, el título del portal puede ser cambiado en el archivo `/xmlui/src/.../message.xml`

4.3.2. Metadatos, *Submission* y *Workflow*

El proceso *Ingest* de DSpace es el principal proceso para realizar depósitos de documentos en DSpace y relaciona a los metadatos, *submission* y *workflow* (ver Fig. 10). El mismo puede iniciarse por aportes externos al sistema o a través de la interfaz de usuario del sistema (XMLUI). Luego el *InProgressSubmission* pasa por varias etapas durante las cuales se le agregan metadatos descriptivos y administrativos y se realizan *checksums* a los contenidos. El resultado de esta secuencia es pasado al *Item Installer*, el cual debe convertir ese objeto en un *Archived Item* de DSpace. El *Item Installer* se encarga de asignar una fecha de adhesión, fecha de disponibilidad, de ser necesario fecha de publicación, procedencia, un identificador persistente, colección a la que pertenece, y crea los índices para su búsqueda y navegación.

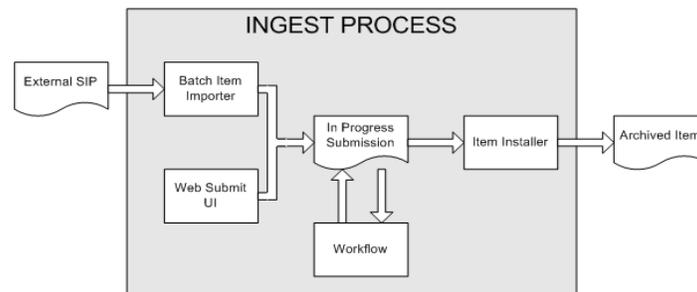


Figura 9. Proceso Ingest. Fuente: Manual de DSpace 1.8.

El formato de metadatos (esquema) a utilizar en DSpace está definido en la sección *Metadata* habilitada para administradores del repositorio en la interfaz web. La recomendación es utilizar formatos estándar con el fin de evitar incertidumbre en el uso de los metadatos, disminuir problemas de interoperabilidad, reducir la variabilidad de criterios, etc. DSpace soporta formatos de metadatos planos (no jerárquicos), por defecto tiene DC calificado. Los formatos de metadatos tentativos a usar son:

- Dublin Core calificado.
- Metadata Object Description Schema (MODS).
- Electronic Theses and Dissertations Metadata Standard (ETDMS).
- Preservation Metadata: Implementation Strategies (PREMIS).
- Europeana Data Model (EDM).

4.3.2.1. Agregar un metadato en DSpace: el primer paso es incorporar el metadato como parte del esquema de metadatos que se usa, para ello, debe ingresar como administrador en el repositorio y en la sección de *Metadata* agregar en el formulario el *element*, *qualifier* y *scope note*. Se pueden realizar las siguientes actividades:

- Habilitar el metadato para ser localizado por una búsqueda. La configuración de los campos habilitados se encuentra en `dspace.cfg`.
- Agregar un campo relacionado con el metadato en el formulario de *submission* a través del archivo `/config/input-forms.xml`.
- Mostrar el metadato según el recurso seleccionado.

4.3.2.2. Proceso de Submission: dividido en las siguientes secciones: preguntas iniciales, formulario de entrada y aceptación de licencias. Dependen de los archivos `/config/item-submission.xml` y `/config/input-forms.xml`. Los pasos tradicionales por defecto para registrar un recurso se observan en la Fig. 11 y dependen de `item-submission.xml`:

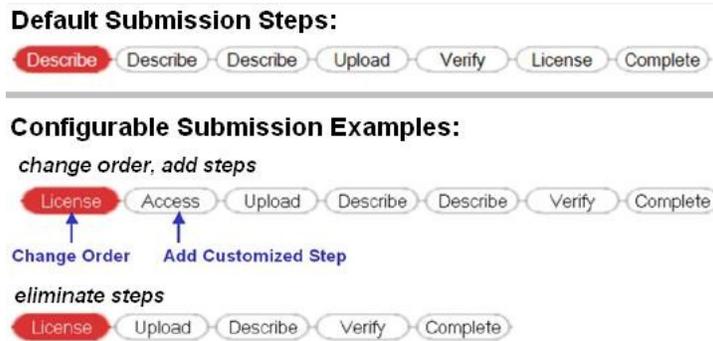


Figura 10. Proceso de Submission

Este proceso puede ajustarse para las diferentes colecciones de DSpace a través de la sección <submission-map> en el archivo de configuración nombrado anteriormente (item-submission.xml) y para cambiar los pasos a seguir, puede hacerse en la sección <step-definitions> y <step> en el mismo archivo.

4.3.2.3. *Flujos de Trabajo (Workflows)*: son procesos opcionales por los que pueden pasar los recursos que han depositado en el repositorio por diferentes vías. Los *workflows* permiten a los administradores de repositorios tener un control de los recursos y sus metadatos antes de estar disponibles formalmente para el público en general. Los flujos de trabajo se asignan a las colecciones y los usuarios pueden recibir notificaciones vía correo electrónico de las diferentes tareas relacionadas. DSpace tiene tres *workflows* que se pueden usar de forma independiente o de forma mixta:

1. Aceptar/Rechazar: este paso se utiliza para permitir a un usuario aceptar o rechazar simplemente el depósito. Si lo rechazan, pueden dar una razón que será enviado por correo electrónico al remitente.
2. Aceptar/Rechazar/Editar Metadatos: este paso se utiliza para permitir a un usuario para aceptar o rechazar el depósito y editar sus metadatos. Si lo rechazan, pueden dar una razón que será enviado por correo electrónico al remitente.
3. Editar Metadatos: este paso se utiliza para permitir al usuario editar los metadatos. Esto podría hacerse para corregir los metadatos o para mejorarla.

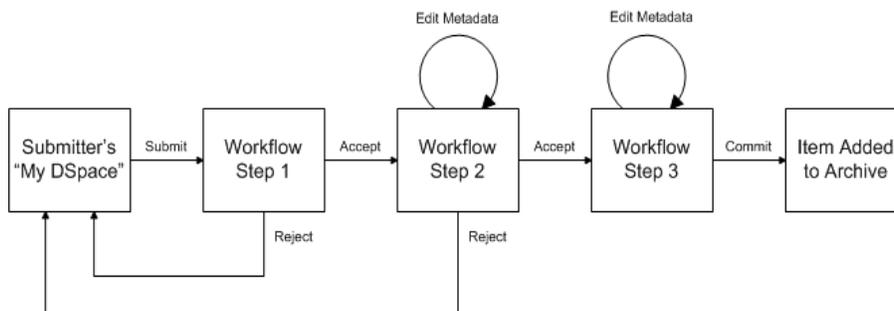


Figura 11. Workflows por Defecto de DSpace.

Los diferentes archivos involucrados en los flujos de trabajo de DSpace son:

- /config/workflow.xml

- /config/workflow.cfg
- /config/xmlui.xconf
- /distribution/dspace-tag-files/.../xml_workflow.sql
- /distribution/dspace-tag-files/.../database_schema.sql
- DescribeStep.java
- DCInput.java

4.3.3. Comunidades y Colecciones

El contenido en el software DSpace está organizado en comunidades que pueden corresponder a entidades administrativas tales como departamentos, laboratorios, oficinas, centros de investigación, entre otros. La comunidad es la categoría más amplia y puede contener subcomunidades y/o colecciones, pero no puede contener recursos. En caso de eliminar la comunidad se eliminarán todas las subcomunidades, colecciones y recursos que contenga. La colección es la categoría donde los recursos son contenidos y no pueden contener otras categorías. Para definir como será la estructura de comunidades y colecciones que se utilizarán, se tiene que tomar en cuenta las siguientes restricciones:

- Las comunidades pueden contener subcomunidades y colecciones.
- Las colecciones solo pueden contener documentos.
- Una colección puede pertenecer solo a una comunidad.
- Los documentos solo pueden estar contenidos en colecciones.
- Un documento puede pertenecer a varias colecciones, aunque solamente una colección es el dueño.

La estructura de comunidades y colecciones recomendada para la propuesta del Repositorio de Documentos Administrativos de la UNET será la siguiente de acuerdo con la distribución de las dependencias:

- RECTORADO

- *Consultoría Jurídica*
- *Centro de Estudios de Telecomunicaciones (CETI)*
- *Consejo de Planificación*
- *Unidad de Desarrollo Físico*
- *Planificación Académica*
- *Auditoría Interna , Unidad de Control*
- *Auditoría Interna , Unidad de Entes Descentralizados*
- *Auditoría Interna , Unidad de Averiguaciones Administrativas*

- VICERRECTORADO ACADÉMICO:

- *Curso Propedéutico*
- *Dirección de Biblioteca*
- *Coordinación de Desarrollo Educativo (CODE)*
- *Coordinación Académica de Desarrollo Curricular*
- *Unidad de Admisión*
- *Programa de Medios Audiovisuales*
- *Programa de Mejoramiento Académico*
- *Programa de Educación a Distancia*
- *Programa de Medios Instruccionales*

- VICERRECTORADO ADMINISTRATIVO

- *Departamento de Computación*
- *Dirección de Finanzas*

- *Departamento de Compras*
- *Departamento de Presupuesto*
- *Departamento de Contabilidad*
- *Departamento de Tesorería*
- *Departamento de Organización y Sistemas*
- SECRETARÍA
 - *Dirección de Comunicación, Información, Protocolo y Relaciones Públicas (DICIPREP)*
 - *Control de Estudios y Evaluación*
 - *Unidad de Control de Estudios*
 - *Unidad de Evaluación*
 - *Coordinación de Asuntos Secretariales*
 - *Unidad de Graduandos*
 - *Unidad de Archivo y Correspondencia*
 - *Unidad de Archivo Histórico*
 - *Unidad de Asuntos Profesorales*
 - *Unidad de Estadística*
- DIRECCIÓN DE RECURSOS HUMANOS
 - *Departamento Administrativo*
 - *Departamento Laboral*
 - *Oficina de Seguros UNET*
- DIRECCIÓN DE SERVICIOS
 - *Departamento de Servicios*
 - *Departamento Administración y Contratos*
 - *Departamento de Mantenimiento*
- DECANATOS
 - *Decanato de Desarrollo Estudiantil*
 - *Coordinación de Bienestar Estudiantil*
 - *Coordinación de Deportes*
 - *Coordinación de Orientación*
 - *Decanato de Postgrado*
 - *Coordinación Académica*
 - *Coordinación Operativa*
 - *Unidad de Administración*
 - *Unidad de Trabajos de Grado*
 - *Unidad de Logística*
 - *Decanato de Extensión*
 - *Coordinación Extensión Sociocultural*
 - *Coordinación de Formación Permanente*
 - *Coordinación de Extensión Agraria*
 - *Coordinación de Extensión Industrial y Pasantías*
 - *Laboratorio de Congelación de Semen - UA La Morusca*
 - *Decanato de Docencia*
 - *Coordinación de Servicio Comunitario*
 - *Departamento de Ingeniería Mecánica*
 - *Núcleo Diseño Mecánico*
 - *Núcleo Materiales y Procesos*
 - *Núcleo Termofluidos*
 - *Núcleo Mecánica del Sólido*
 - *Departamento Ingeniería Industrial*
 - *Núcleo Gerencia Organizacional*
 - *Núcleo de Producción*
 - *Núcleo de Técnicas Cuantitativas*
 - *Departamento de Ingeniería Ambiental*

- - Núcleo Ingeniería Ambiental
- - Núcleo Procesos Ambientales
- *Departamento de Ingeniería Agronómica*
 - *Núcleo de Ecología*
 - *Núcleo de Ingeniería*
 - *Núcleo de Desarrollo Agrario*
 - *Núcleo de Producción y Procesamiento*
- *Departamento Ingeniería de Producción Animal*
 - *Núcleo I Fisiología y Metabolismo Animal*
 - *Núcleo II Producción Animal Aplicada*
 - *Núcleo III Tecnología y Salud Animal*
 - *Núcleo IV Apoyo a la Producción Animal*
- *Departamento de Ingeniería Electrónica*
 - *Núcleo de Electricidad*
 - *Núcleo de Electrónica y Sistemas Digitales*
 - *Núcleo de Telecomunicaciones*
 - *Núcleo de Instrumentación, Control y Señales*
- *Departamento de Arquitectura*
 - *Núcleo 1 de Proyectos*
 - *Núcleo 2 de Proyectos*
 - *Núcleo Producción y Tecnología*
 - *Núcleo de Contexto Ambiental*
 - *Núcleo de Historia de la Arquitectura*
 - *Núcleo Sistemas de Representación y Simulación*
- *Departamento de Ingeniería Informática*
 - *Núcleo Tecnología Básica*
 - *Núcleo de Telemática*
 - *Núcleo de Ingeniería de la Información*
- *Departamento de Matemática y Física*
 - *Núcleo Cálculo*
 - *Núcleo Matemática*
 - *Núcleo Matemática Aplicada*
 - *Núcleo de Física*
- *Departamento de Química y Biología*
 - *Núcleo I Química Básica*
 - *Núcleo II Química Avanzada*
- *Departamento de Ciencias Sociales*
 - *Núcleo Idiomas*
 - *Núcleo Desarrollo Humano*
 - *Núcleo Estudios Sociales y Económicos*
- *Coordinación de Laboratorios y Proyectos*
 - *Laboratorio de Computación Valerio Wong*
 - *Laboratorio de Química y Ciencias del Agro*
 - *Laboratorio de Química I*
 - *Laboratorio de Electrónica*
 - *Laboratorio de Ingeniería*
 - *Lab. de Termofluidos, Termodinámica II, Plantas de Potencia y Máquinas de Fluidos*
 - *Lab. Procesos de Manufactura I*
 - *Lab. Electrotecnia I e Ing. Eléctrica*
 - *Laboratorio de Máquinas y Herramientas*
 - *Laboratorio Bioambiental*
- *Departamento Ingeniería Civil*
 - *Núcleo de Vialidad*

- *Departamento Licenciatura en Música*
 - *Núcleo Pedagógico*
 - *Núcleo Práctica y Ejecución Instrumental Musical*
 - *Núcleo Lenguaje Musical*
- *Departamento Entrenamiento Deportivo*
 - *Núcleo Gestión Deportiva*
 - *Núcleo Prácticas Deportivas*
 - *Núcleo Bases del Entrenamiento Deportivo*
- *Departamento de Ciencias de la Salud*
- *Carrera Ing. Agroindustrial*
- *Departamento de Carreras Técnicas Semipresenciales*
- *Oficinas Académicas del TSU en Agropecuaria*
- *Oficinas Académicas El Piñal*
- *Oficinas Académicas Pregonero*
- *Oficinas Académicas La Grita*
- *Unidades de Producción*
 - *La Tuquerena*
 - *Planta Productos Lácteos*
 - *Planta Productos Cárnicos*
 - *Lab. De Congelación de Sémén*
 - *Planta Productos Vegetales*
- *Unidades Académicas*
 - *Núcleo Rural La Pradera*
 - *Núcleo El Piñal*
 - *La Primavera*
 - *La Tuquerena*
- *Decanato de Investigación*
 - *Coordinación de Investigación Agropecuaria*
 - *Coordinación de Investigación Industrial*
 - *Coordinación de Inv. Socio-Económico-Cultural*
 - *Coordinación de Inv. Ciencias Exactas y Naturales*
 - *Unidad de Proyectos Especiales*
 - *Laboratorio de Catálisis*
 - *Núcleo de Investigación Apícola*
 - *Centro de Inv. en Control Biológico*
 - *Jardín Botánico del Táchira*
 - *Laboratorio de Fitoquímica*
 - *Laboratorio de Inv. en Mejoramiento Animal*
 - *Laboratorio de Biofertilizantes*
 - *Responsable del GIAS*
 - *Responsable Académico de la Estación Experimental de Piscicultura*
 - *Jefe Centro de Estudio de Vectores*
 - *Jefe de Laboratorio de Investigaciones Genéticas*
- **OFICINA DE ENLACE**
- **CONSEJO SUPERIOR**
- **ASOCIACIONES**
 - *APUNET, Asociación de Profesores de la UNET*
 - *AEAUNET, Asociación de Empleados Administrativos*
 - *SUTUNET, Sindicato Único de Trabajadores de la UNET*
 - *SIPROUNET, Sindicato de Profesores de la UNET*

Entonces, las comunidades principales están en mayúsculas y en negritas, por ejemplo: Rectorado, Vicerrectorado Académico, etc. Las subcomunidades están en cursivas y en

negritas, y las colecciones son todas las que no están en negritas. Luego, se realiza una importación a DSpace vía consola, con todas las comunidades y colecciones a través de un archivo XML con la siguiente estructura:

```

<import_structure>
  <community>
    <name>Community Name</name>
    <description>Descriptive text</description>
    <intro>Introductory text</intro>
    <copyright>Special copyright notice</copyright>
    <sidebar>Sidebar text</sidebar>
    <community>
      <name>Sub Community Name</name>
      <community> ...[ad infinitum]...
    </community>
  </community>
  <collection>
    <name>Collection Name</name>
    <description>Descriptive text</description>
    <intro>Introductory text</intro>
    <copyright>Special copyright notice</copyright>
    <sidebar>Sidebar text</sidebar>
    <license>Special licence</license>
    <provenance>Provenance information</provenance>
  </collection>
</import_structure>

```

Tabla 4. Formato XML para Importar Comunidades - Colecciones

4.3.4. Control de Autoridades

El control de autoridad es el mecanismo que permite crear y mantener términos en algún índice relacionado al material bibliográfico de un catálogo documental. Cumple dos funciones importantes:

- Permite a los catalogadores eliminar la ambigüedad de artículos con títulos similares o idénticos. Por ejemplo, dos autores que por casualidad han publicado bajo el mismo nombre se pueden distinguir unos de otros mediante la adición de la inicial del segundo, fechas de nacimiento y/o muerte.
- Permite mantener materiales lógicamente equivalentes aunque presentados de manera diferente. Por ejemplo, los registros de autoridad se utilizan para establecer los títulos uniformes, que puede colocar todas.

A continuación se incluye una serie de metadatos para los que se debería utilizar mecanismos de control de autoridades. Se agrupan, según su grado de variabilidad, en dos tipos:

1. *Autoridades estables*: dada la baja frecuencia con la que estos vocabularios son modificados por agregado, edición o remoción de términos. En estos casos, es recomendable pensar en realizar un mantenimiento manual, directamente sobre los datos (bases de datos, XML, o similar) y así evitar la implementación de módulos administrativos:
 - Tipos de documento

- Idiomas
 - Sistemas de clasificación
 - Lista de encabezamiento de Materias
 - PIDU (País - Institución - Dependencia - Unidad)
 - Grados alcanzados
 - Frecuencias
 - Indizadores de revistas
 - Licencias
 - Tesoros
2. *Autoridades volátiles*: son vocabularios con alta frecuencia de actualización. Es deseable contar con alguna herramienta de ABM (alta, baja y/o modificación) para su administración por parte de los administradores:
- Autores
 - Revistas
 - Identificadores geográficos
 - Eventos: congresos, seminarios y jornadas
 - Títulos de revistas

4.3.4.1. Control de autoridades en DSpace

Depende principalmente de dos archivos, `dspace.cfg` y `input-forms.xml`. En el archivo `dspace.cfg` se encuentra todos los parámetros de configuración, tales como: *plugins*, modalidades del control de autoridades para cada metadato, características, etc. En `input-forms.xml` se encuentra la definición del proceso de *submission* y se encuentran los elementos *value-pairs* que permiten restringir algunos tipos de metadatos. DSpace ofrece dos funcionalidades por separado, que pueden ser configurados y utilizados:

- *Choice management mechanism*: es el mecanismo para selección o elección de los posibles valores de un metadato a partir de valores propuestos o consultas. Estas pueden ser cerradas (*false*) o abiertas (*true*).
- *Authority control*: base de datos que contiene el vocabulario controlado y puede configurarse como opcional.

Ambas funcionalidades se pueden aplicar a cualquier metadato. Para comprender con mayor profundidad el concepto de control de autoridades de acuerdo con las diferentes funcionalidades que ofrece DSpace, se tienen las siguientes definiciones:

- *Autoridad*: conjunto de valores fijos identificado por una clave.
- *Registro de autoridad*: información asociada con uno de los valores de la autoridad.
- *Clave de autoridad*: un identificador persistente que se corresponde con el registro de autoridad.

DSpace ofrece las siguientes clases e interfaces, que pertenecen al paquete `org.dspace.content.authority` para el control de autoridades:

- *Choice*, clase que contiene los atributos *authority*, *label* y *value*.
- *Choices*, clase que contiene un conjunto de *Choice*.

- `ChoiceAuthority`, interfaz para suplir el mecanismo del control de autoridades.

4.3.4.1.1. La clase `org.dspace.content.authority.Choice`: representa el contenido de una elección referente a uno de los términos controlados por las autoridad de nombre. Las instancias de esta clase deben ser creadas para resolver los requerimientos de términos controlados. Esta clase no presenta comportamientos, sino métodos para acceder a sus variables y un solo constructor que inicializa sus variables, y está compuesto por los siguientes atributos:

- `String label`, es la descripción del la clave de autoridad.
- `String authority`, es la clave de autoridad dentro de la autoridad de nombre, se espera que este "opaco" y persistente que identifique unívocamente un registro de autoridad.
- `String value`, es el texto a insertar dentro del campo de metadatos del ítem en cuestión.

4.3.4.1.2. La clase `org.dspace.content.authority.Choices`, contiene un conjunto (arreglo) de `Choice` para ser devuelto por una autoridad en respuesta a una búsqueda. Con lo cual hay que tener presente elegir un valor de confianza adecuado, para el conjunto de `Choice` que serán incluidos en la búsqueda.

4.3.4.1.3. La interfaz `org.dspace.content.authority.ChoiceAuthority` para implementar *plugins* sobre el control de autoridad. Esta interfaz se define para tomar valores de una autoridad que puede ser provista de cualquier fuente ya sea una base de datos local o no, un servicio prestado por un tercero, documentos en cualquier formato y algo que provea términos controlados. La interfaz cuenta con tres métodos:

- Método `getMatches`. La idea de este método es que el `ChoiceAuthority` tome todos los valores de una autoridad, que coincidan con un valor provisto por el usuario. Este método es usado para retornar el conjunto de los términos controlados incluidos en un objeto de la clase `Choices` y nunca debe ser nulo.
- Método `getBestMatch`. Este método se espera que retorne el mejor término controlado encontrado si existe, incluidos en un `Choices`. Este método se usa típicamente en ingresos de metadatos en modo no interactivo (en modo *batch*) donde no hay un agente interactivo para elegir entre las opciones.
- Método `getLabel`. Este método se utiliza para obtener una etiqueta (*label*) visible para el usuario de una clave de autoridad, correspondiente a un término controlado. Esto puede ser llamado varias veces al rellenar una página web por lo que deben aplicarse lo más eficientemente posible.

Los diferentes tipos de controles de valores para metadatos posibles en DSpace son:

- Valores en el `input-form.xml`
- Valores definidos en una clase propia
- Valores definidos en un XML
- Valores definidos fuera del servidor

Independientemente del tipo de control de autoridad a aplicar, el DSpace no tiene activado

(por default) el control de autoridad. Para habilitarlo hay que modificar el archivo de configuración [Dspace-DIR]/config/dspace.cfg agregando o modificando la siguiente línea: `webui.controlledvocabulary.enable = true`.

4.3.5. Tipos de Documentos

Las dependencias del Rectorado, Secretaría y el Departamento de Ingeniería en Informática de la UNET, aportaron diferentes documentos generados desde su creación, los cuales se analizaron para definir los tipos de documentos recomendados para el repositorio, a saber:

1. Acta
2. Constancia
3. Memorando
4. Oficio
5. Resolución
6. Jurídico

A continuación una definición de los tipos de documentos y una lista de los metadatos asociados a cada uno de ellos:

4.3.5.1. Acta: relación escrita de lo sucedido, tratado o acordado. Los metadatos que identifican este tipo de documento son:

- *Tipo de Doc*
- *Fulltext*
- *ID*
- *Fecha*
- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Persona a certificar*
- *Dependencia de la persona a certificar*

4.3.5.2. Constancia: escrito en que se hace constar algún acto o hecho. Los metadatos que lo identifican son:

- *Tipo de Doc*
- *Fulltext*
- *ID*

- *Fecha*
- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Persona a certificar*
- *Dependencia de la persona a certificar*

4.3.5.3. *Memorando*: alguna situación que debe recordarse. Sus metadatos son:

- *Tipo de Doc*
- *Fulltext*
- *ID*
- *Fecha*
- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Persona que recibe*
- *Dependencia que recibe*
- *Personas en el texto*
- *Dependencias en el texto*

4.3.5.4. *Oficio*: comunicación escrita, referente a los asuntos de las administraciones públicas. Los metadatos correspondientes son:

- *Tipo de Doc*
- *Fulltext*
- *ID*
- *Fecha*

- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Persona que recibe*
- *Dependencia que recibe*
- *Personas en el texto*
- *Dependencias en el texto*

4.3.5.5. *Resolución*: decreto o fallo de autoridad gubernativa o judicial. Los diferentes metadatos que se extraen de este tipo de documento son:

- *Tipo de Doc*
- *Fulltext*
- *ID*
- *Fecha*
- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Tipo del Consejo*
- *Tipo de Sesión*
- *Personas en el texto*
- *Dependencias en el texto*
- *Puntos de Agenda*
- *Docs relacionados*

4.3.5.6. *Jurídico*: para esta clasificación se tienen varios subtipos:

- *Ley*: en el régimen constitucional, disposición votada por las Cortes y sancionada

por el Jefe del Estado.

- Reglamento: colección ordenada de reglas o preceptos, que por la autoridad competente se da para la ejecución de una ley.
- Norma: precepto jurídico. Regla que se debe seguir o a que se deben ajustar las conductas, tareas, actividades, etc.
- Decreto: decisión de un gobernante o de una autoridad, o de un tribunal o juez, sobre la materia o negocio en que tengan competencia. También es conocido como un dictamen.
- Gaceta: publicación periódica en la que se dan noticias comerciales, administrativas, literarias o de otra índole.
- Instructivo: que instruye o sirve para instruir, es decir, comunicar sistemáticamente ideas, conocimientos o doctrinas.

Los metadatos para todos los tipos de documentos jurídicos son:

- *Tipo de Doc*
- *Fulltext*
- *ID*
- *Fecha*
- *Ciudad*
- *Localización*
- *Título/Asunto*
- *Título Alternativo*
- *Persona suscriben*
- *Dependencia suscriben*
- *Iniciales de Supervisores*
- *Quien transcribió doc*
- *Varios*
- *Palabras Clave*
- *Otros descriptores*
- *Número de Artículos*
- *Páginas*
- *Docs relacionados*

Finalmente, se elaboraron dos tablas para resumir los diferentes metadatos que estarán presentes en las dos páginas (vía Web) del proceso de *submission*. Los metadatos son configurables en el archivo `workflow.xml`:

PAGINA 1					
Label	Metadato	Obligat.	Repet.	Controlado	Scope
Tipo de Doc	dc.type	SI	NO	VALUE-PAIRS	ALL
Fulltext	rda.fulltext	SI	NO	VALUE-PAIRS	WORKFLOW
ID	rda.id	SI	NO		ALL
Fecha	dc.date.issued	SI	NO	NO	ALL
Ciudad	rda.ciudad	SI	NO	BBDD	ALL
Localizacion	mods.location	NO	SI	NO	ALL
Titulo/Asunto	dc.title	SI	NO	NO	ALL
Título Alternativo	dc.title.alternative	NO	SI	NO	WORKFLOW
Persona suscriben	rda.contributor.director	SI	SI	BBDD	ALL
Dependencia suscriben	mods.originInfo.place	SI	NO	BBDD	ALL
Iniciales de Supervisores	rda.iniciales	NO	SI	NO	ALL
Quien transcribio doc	rda.creator.transcriptor	SI	NO	BBDD	ALL

Tabla 5. Primera Página del Proceso Submission

PAGINA 2						
Label	Metadato	OB.	RE.	Control.	Scope	TYPE-BIND
Varios	dc.description.abstract	NO	SI	NO	ALL	
Palabras Clave	rda.subject.keyword	NO	SI	NO	ALL	
Otros descriptores	rda.subject.other	NO	SI	PROPIO	WORKFLOW	
Persona a certificar	rda.creator.certificar	NO	SI	BBDD	ALL	Acta/Constancia
Depen. de la pers. a certificar	rda.creator	NO	SI	BBDD	ALL	Acta/Constancia
Persona que recibe	rda.recibe.persona	NO	NO	BBDD	ALL	Memorando/Oficio
Dependencia que recibe	rda.recibe.dependencia	NO	NO	BBDD	ALL	Memorando
Personas en el texto	rda.texto.persona	NO	SI	BBDD	ALL	Memorando/Resolucion
Dependencias en el texto	rda.texto.dependencia	NO	SI	BBDD	ALL	Memorando/Resolucion
Numero de Articulos	rda.numarticulo	NO	NO	NO	ALL	Juridicos
Páginas	rda.pagina	NO	NO	NO	ALL	Juridicos
Docs relacionados	rda.relation	NO	SI	NO	ALL	Juridicos/Resolucion
Tipo del Consejo	rda.tipoconsejo	NO	NO	NO	ALL	Resolucion
Tipo de Sesión	rda.sesion	NO	NO	NO	ALL	Resolucion
Puntos de Agenda	rda.temas	NO	SI	NO	ALL	Resolucion

Tabla 6. Segunda Página del Proceso Submission

4.3.6. OAI-PMH

DSpace soporta el *Open Archive Initiative Protocol for Metadata Harvesting* (OAI-PMH) en su versión 2.0 como un proveedor de datos. Este protocolo establece un conjunto de reglas a partir de las cuales se puede realizar un intercambio de recursos de forma confiable y exitosa. Este servicio es desplegado como otra aplicación web de DSpace, por ello es independiente de los otros WebUI y se ubica en la capa de aplicación. Este servicio puede ser usado a través del URL de la instalación realizada, `http://[full-URL-to-OAI-PMH]/request?verb=Identify`, y la aplicación web se puede encontrar en `/[dspace]/webapps/oai/`. El servicio está configurado por defecto, pero si se desean realizar ajustes se tiene que acceder al archivo `/[dspace.dir]/install/config/modules/oai.cfg`. Adicionalmente, se tiene que crear un enlace en la carpeta *webapps* del Apache Tomcat, apuntando al OAI de la instalación del

DSpace-Sedici, de la siguiente forma: `ln -s /home/.../[dspace.dir]/install/webapps/oai/ oai`. También, se puede visitar el OAI de la versión demo de DSpace, <http://demo.dspace.org/oai/request?verb=Identify>.

4.3.7. Discovery

Es un módulo que funciona en la interfaz de usuario XML de DSpace, que permite habilitar los *facets* para búsqueda y navegación dentro del repositorio, sin necesidad de estar registrado. Para poner en marcha el *Discovery* en DSpace 1.8. se deben seguir los siguientes pasos:

- Verificar que las rutas de la versión de producción de DSpace-Sedici (`[dspace.dir]/install/`) estén correctas en los archivos `server.xml` y `dspace.xml` que se encuentran en la carpeta del Apache Tomcat configurado en Eclipse.
- Crear un enlace en la carpeta *webapps* del Apache Tomcat, apuntando al SOLR de la instalación del DSpace-Sedici, de la siguiente forma:
`ln -s /home/.../[dspace.dir]/install/webapps/solr/ solr`
- Verificar ruta del SOLR en `discovery.cfg`, ubicado en `[dspace.dir]/install/config/modules/`, ya que en `dspace.cfg` y `xmlui.xconf` se tienen la configuración mínima para el funcionamiento del *Discovery*.
- Indexar los recursos que ya tiene el DSpace-Sedici, a partir del directorio `bin` del *install* (`[dspace.dir]/install/bin/`), de la siguiente manera:
`./dspace update-discovery-index -f`
- El mensaje que se debe generar si todo está configurado correctamente es el siguiente:

```
SLF4J: The requested version 1.5.8 by your slf4j binding is not compatible with [1.5.5, 1.5.6].
SLF4J: See http://www.slf4j.org/codes.html#version_mismatch for further details.
```

4.3.8. Usuarios y Grupos

Aunque muchas de las funciones de Dspace tales como el navegación y recuperación de documentos se puede hacer en forma anónima, algunas características (y tal vez algunos documentos) sólo están disponibles para determinados usuarios "privilegiados". En DSpace existen los *e-person* y *groups*, que son la manera de identificar a los usuarios para la concesión de privilegios. Esta identidad está ligada a una sesión de una aplicación DSpace tales como la interfaz de usuario Web o uno de los programas por línea de comandos.

4.3.8.1. *E-Person*: existen dos formas de agregar usuarios nuevos, una desde la administración del repositorio y la otra a través de una solicitud por parte del usuario. El administrador es el único que puede conceder privilegios a los usuarios dentro de las categorías. A continuación se muestran las diferentes características que DSpace necesita de cada *e-person*:

- Nombres y apellidos
- Correo electrónico.
- Una contraseña encriptada.
- Una lista de las colecciones del *e-person* de las cuales desea ser notificado de

nuevos recursos.

- El usuario puede ser autenticado vía LDAP.

4.3.8.2. *Groups*: sirven para gestionar a los usuarios y de esa manera ellos pueden tener privilegios sobre comunidades, subcomunidades y/o colecciones. DSpace tiene por defecto tres grupos: *Administrator* que tiene todos los privilegios sobre el repositorio, *Anonymous* que permite acceder a cualquier recurso del repositorio por eso no se necesita estar registrado y *Summit* es el grupo para autorizar nuevos recursos a la colección. Adicionalmente, es importante destacar que se pueden crear la cantidad de grupos que se deseen. A continuación se listan las acciones de autorización que puede ser aplicarse:

- En colecciones: ADD/REMOVE, DEFAULT_ITEM_READ, DEFAULT_BITSTREAM_READ y COLLECTION_ADMIN.
- En items: ADD/REMOVE, READ y WRITE.
- En bundles: ADD/REMOVE.
- En bitstreams: READ y WRITE.

4.4. Plataforma de Trabajo

4.4.1. Gestor de Proyectos

La recomendación es usar un gestor de proyectos que permita tener un control de las versiones del software, tickets de mantenimiento y errores presentados, actividades, etc. La recomendación es usar Redmine 2.1.2 y un sistema de control de versiones llamado Apache Subversion 1.6.19. Es importante destacar que DSpace 3 estará integrado con el sistema de control de versiones GIT. En la siguiente figura se muestra el gestor de proyectos Redmine que usa SeDiCI:

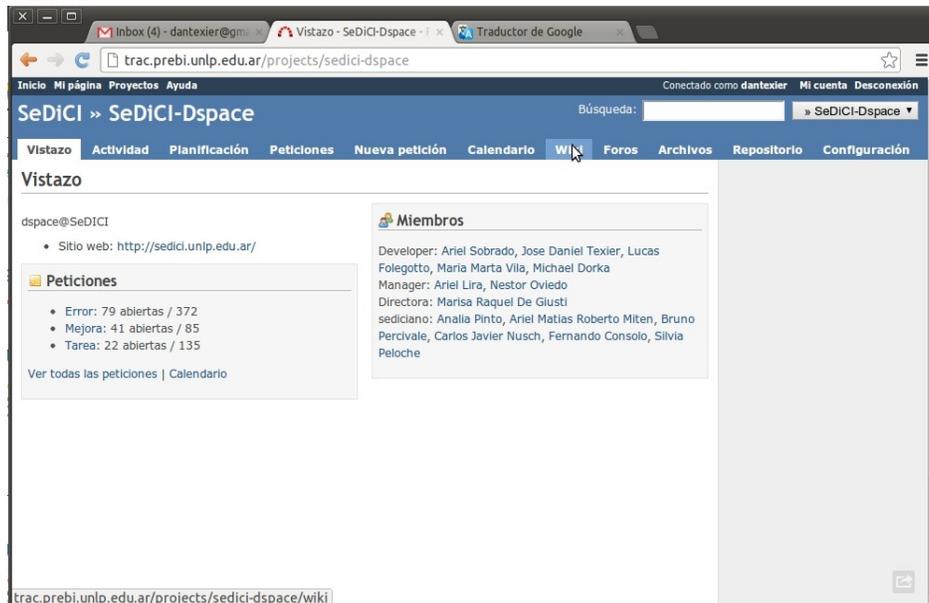


Figura 12. Redmine de SeDiCI

4.4.2. Entorno de Desarrollo Integrado

En secciones anteriores, se ha explicado la necesidad de poder enlazar un entorno de

desarrollo de software con un sistema de control de versiones. Por eso la recomendación es Eclipse Juno como herramienta de programación y Apache Subversion como sistema de control de versiones. En la figura 14, se muestra el entorno de desarrollo integrado funcionando:

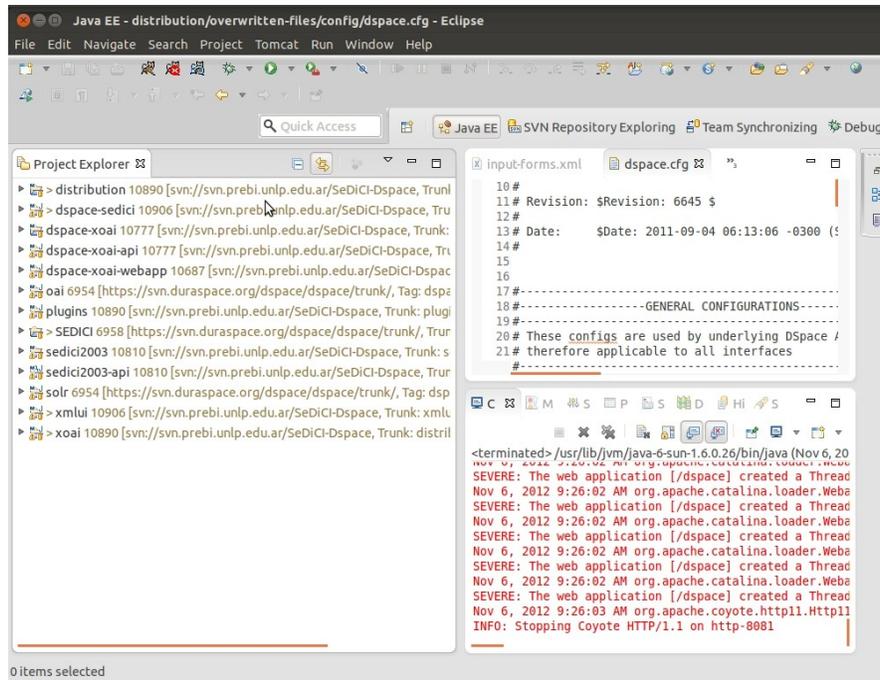


Figura 13. Entorno de Desarrollo Integrado

4.4.3. Servidores

La recomendación para un proyecto de desarrollo de software es definir una política de tres servidores: uno de desarrollo, uno de pruebas y uno de producción. Para este proyecto, el servidor de desarrollo estará en cada computadora de los programadores y sincronizados con el control de versiones. Luego, el servidor de pruebas estará sincronizado con la carpeta *target* y finalmente, el servidor de producción se enlazará con la carpeta */install/webapps/* del proyecto DSpace.

4.6. Catalogación

Este proceso estará controlado por los bibliotecarios del repositorio, quienes aprobarán cada uno de los depósitos de recursos realizados por las diferentes vías permitidas, tales como: autoarchivo por parte de los usuarios, depósitos por parte de catalogadores y depósitos automáticos.

4.6.1. Autoarchivo

En este proceso los autores tienen que introducir los documentos y los metadatos que describen a los recursos a través del portal del repositorio. Una vez finalizado debe aprobarse por el grupo bibliotecarios habilitados de acuerdo con las políticas de cada repositorio. Esta aprobación es necesaria para garantizar la no incorporación de contenidos

inapropiados. Este autoarchivo estaría habilitado por al menos un usuario de cada dependencia autorizada por la Universidad.

4.6.2. Depósitos por Catalogadores Propios

Una de las políticas de los repositorios es depositar recursos que pueden ser recolectados por diferentes vías, una vez detectados esos recursos los catalogadores (bibliotecarios) van ingresando los metadatos correspondientes para cada uno de los recursos encontrados. Los depósitos realizados por bibliotecarios del repositorio no necesitan de aprobación y por ende, son habilitados inmediatamente para ser mostrados al público en el portal.

4.6.3. Depósitos Automáticos

La propuesta está diseñada para depositar los diferentes recursos generados por la Universidad, pero es importante ir recuperando todos los documentos generados por las dependencias que la política institucional regule. Por ello, se deben diseñar procesos de recuperación de todos esos documentos e implementarlos por programadores, ya que es necesario extraer la mayor cantidad de metadatos con una calidad adecuada, para luego ser aprobados, y algunas veces modificados, por los bibliotecarios.

4.7. Documentación

La propuesta ayuda a lograr una configuración básica de acuerdo con la normativa legal de la República Bolivariana de Venezuela para gestionar los documentos administrativos de una Universidad Venezolana. En caso que se desee profundizar en la configuración de DSpace, se recomiendan estos enlaces:

- Página oficial de DSpace, <http://www.dspace.org>
- Manual oficial de DSpace: https://github.com/DSpace/DSpace/raw/dspace-1_8_x/dspace/docs/DSpace-Manual.pdf
- Wiki oficial de DSpace: <https://wiki.duraspace.org/>

Capítulo 5

Consideraciones Finales

5.1. Recomendaciones

1. Una vez implementado el REDAUNET, se recomienda enlazarlo con los usuarios de la Intranet de la UNET vía LDAP, ya que de esa manera DSpace trabajaría con las cuentas existentes de la UNET y no guardaría las claves de identificación. De igual manera se recomienda controles de seguridad según las direcciones IP de las computadoras de las dependencias administrativas y académicas.
2. Para las comunidades y colecciones se pueden diseñar flujos de trabajo, que permitan controlar de forma distribuida los diferentes recursos que se depositan, a través de autorizaciones por las comunidades y colecciones deseadas.
3. Establecer reuniones de trabajo con organismos superiores de la UNET, como el Ministerio del Poder Popular para la Educación Universitaria y la Oficina de Planificación del Sector Universitario, para definir procedimientos de interoperabilidad de los diferentes recursos catalogados que permitan controlar y supervisar procesos como: presupuesto, matrícula estudiantil, becas, salarios, cumplimiento de calendarios académicos, etc.
4. DSpace ofrece módulos que permiten establecer políticas de preservación y difusión de los recursos, por ello, se recomienda diseñar procedimientos para implementarlos en el repositorio garantizando el acceso para generaciones futuras o para posibles auditorías.
5. DSpace está en constante evolución y aproximadamente cada año pone a disposición versiones nuevas de su producto, por tanto, se recomienda mantener el proyecto configurado para estar actualizados y evitar trabajos mayores para estar al día con las versiones ofrecidas.

5.2. Conclusiones

1. Este trabajo contribuye un aporte al área de repositorios digitales, sobre todo, en el ámbito administrativo y es un modelo para que las instituciones gestionen y preserven sus documentos.
2. Se realizó un relevamiento bibliográfico del dominio de los Repositorios de Documentos Administrativos, partiendo del concepto de las Bibliotecas Digitales y Repositorios Institucionales, y terminando con algunos estándares y tecnologías que se están aplicando en este dominio. Este relevamiento permitió confirmar la necesidad de un repositorio de documentos y deja en evidencia que el tema está en estudio y en constante evolución.
3. La UNET cuenta con diversos tipos de documentos administrativos, pero para la propuesta se definieron los siguientes tipos: actas, constancias, memorandos, oficios, resoluciones y jurídicos que está subdividido en: leyes, reglamentos, decretos, normas, etc. Luego de este análisis, se definieron los distintos metadatos que permitirán identificar y localizar los documentos correctamente.
4. En el trabajo realizado se observó que las instituciones venezolanas cuentan con una

legislación coherente con los principios del acceso abierto pero requieren iniciativas que promuevan el uso de repositorios que garanticen la publicación y preservación de los documentos generados.

5. Se generó la documentación para la instalación, mantenimiento y desarrollo de una plataforma de software para el REDAUNET, disponible vía Web.
6. Es necesario explorar qué sucede en otras instituciones con los documentos administrativos mediante cuestionarios similares al aplicado en este trabajo, porque se podrían hacer propuestas cónsonas con las expectativas institucionales que permitan la catalogación y preservación de los documentos administrativos.

5.3. Trabajos Futuros

1. Diseñar una estrategia de cursos en línea sobre derechos de autor y la filosofía del Open Access, para lograr generar una masa crítica sobre la necesidad de los repositorios para la universidad, además de lograr involucrar a la institución dentro del marco legal reglamentado en el país.
2. La arquitectura de DSpace permite diseñar interfaces sin afectar la lógica del repositorio, tales como JPSUI (interfaz con soporte la tecnología JavaServer Pages) o XMLUI (interfaz soportada por Framework Apache Cocoon), por ello, se recomienda diseñar interfaces para personas discapacitadas de acuerdo con la Ley para Las Personas con Discapacidad de la República Bolivariana de Venezuela promulgada en el 2007.
3. Desarrollar una plataforma de gestión y creación de archivos en línea, que a su vez permita la catalogación automática en el repositorio.
4. Realizar un manual de referencia del proceso de migración entre las versiones de DSpace para mantener actualizada la plataforma del REDAUNET.
5. La propuesta implementada se realizó en DSpace 1.8, por ello, se recomienda hacer una actualización a la versión siguiente de DSpace (3.1). Los cambios realizados entre versiones se pueden observar en el roadmap de DSpace.

Referencias

- [1] M. Castells, *The Rise of the Network Society: The Information Age: Economy, Society, and Culture Volume I*, 2nd Edition with a New Preface. Wiley-Blackwell, 2009.
- [2] J. De Souza Silva, J. Cheaz Peláez, y J. Calderón Romero, *La cuestión institucional, de la vulnerabilidad a la sostenibilidad institucional en el contexto del Cambio de Epoca*. Costa Rica: Servicio Internacional para la Investigación Agrícola Nacional - ISNAR, 2001.
- [3] M. Prensky, "Digital Natives, Digital Immigrants," *MCB University Press*, vol. 9, no. 5, Oct. 2001.
- [4] comScore, "comScore, Inc. - Measuring the Digital World," 2013. [En línea]. Disponible: <http://www.comscore.com/>. [Última visita: 21-Feb-2013].
- [5] OpenDOAR, "OpenDOAR - Home Page - Directory of Open Access Repositories," 2013. [En línea]. Disponible: <http://www.opendoar.org/>. [Última visita: 21-Mar-2013].
- [6] ROAR, "Registry of Open Access Repositories (ROAR)," 2013. [En línea]. Disponible: <http://roar.eprints.org/>. [Última visita: 21-Mar-2013].
- [7] L. Domènech, J. M. Barrueco Cruz, A. Caballos Villar, Á. Campos Rodríguez, N. Casaldàliga, P. Combarro Felpeto, R. Cívico Martín, M. A. García Gil, M. Losada, J. C. Morillo Moreno, y S. Reoyo, "Guía para la evaluación de repositorios institucionales de investigación," 2010. [En línea]. Disponible: <http://digital.csic.es/handle/10261/35735>. [Última visita: 30-Ago-2012].
- [8] UNET, "UNET - Secretaría," 2013. [En línea]. Disponible: <http://secretaria.unet.edu.ve/>. [Última visita: 30-Mar-2013].
- [9] X. Agenjo y F. Hernández, "Tendencias internacionales en el desarrollo funcional de la recuperación de la información: Linked Open Data (LOD)," X Workshop Rebiun sobre proyectos digitales: diez años de proyectos digitales: cambian las bibliotecas, cambian los profesionales. Valencia, 7 y 8 de octubre de 2010.
- [10] S. Nguyen y G. Chowdhury, "Digital Library Research (1990-2010): A Knowledge Map of Core Topics and Subtopics," in *Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation*, vol. 7008, C. Xing, F. Crestani, and A. Rauber, Eds. Springer Berlin / Heidelberg, 2011, pp. 367–371.
- [11] G. G. Chowdhury y S. Chowdhury, "Digital library research: major issues and trends," *Journal of Documentation*, vol. 55, no. 4, pp. 409–448, Ene. 1999.
- [12] S. Yang, J. Pomerantz, B. M. Wildemuth, y E. A. Fox, "Curriculum development for digital libraries," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, 2006. JCDL '06*, 2006, pp. 175–184.
- [13] C. L. Liew, "Digital library research 1997-2007: Organisational and people issues," *Journal of Documentation*, vol. 65, no. 2, pp. 245–266, Jun. 2009.
- [14] SJC, "Scimago Journal & Country Rank," 2013. [En línea]. Disponible: <http://www.scimagojr.com/>. [Última visita: 28-Mar-2013].
- [15] S. M. Griffin, "Taking the Initiative for Digital Libraries.," *Electronic Library*, vol. 16, no. 1, pp. 24–27, 1998.
- [16] DLI, "Digital Libraries Initiative - Phase II," 1998. [En línea]. Disponible:

- <http://www.nsf.gov/pubs/1998/nsf9863/nsf9863.htm>. [Última visita: 09-May-2012].
- [17] M. Lesk, *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann, 1997.
- [18] D. Waters, "What Are Digital Libraries?," *CLIR Issues*, vol. 4, 1998.
- [19] C. L. Borgman, "What Are Digital Libraries? Competing Visions.," *Information Processing & Management*, vol. 35, no. 3, pp. 227–243, 1999.
- [20] D. Bawden, I. Rowlands, y B. L. R. y I. Centre, *Understanding Digital Libraries: Towards a Conceptual Framework*. British Library Research and Innovation Centre, 1999.
- [21] T. R. Kochtanek y K. K. Hein, "Delphi study of digital libraries," *Information Processing & Management*, vol. 35, no. 3, pp. 245–254, May 1999.
- [22] W. Y. Arms, *Digital Libraries*. MIT Press, 2001.
- [23] T. Sharon y A. Frank, "Digital libraries on the Internet," *66th IFLA Council and General Conference*, 2000.
- [24] J. Tramullas Saz, "Propuestas de concepto y definición de la biblioteca digital," pp. 11–20, 2002.
- [25] I. H. Witten, D. Bainbridge, y D. M. Nichols, *How to Build a Digital Library, Second Edition*, 2nd ed. Morgan Kaufmann, 2009.
- [26] J. Tramullas, "Bibliotecas digitales," VI Seminario de Centros de Documentación Ambiental y Espacios Naturales Protegidos, 2007.
- [27] Leonardo Candela, Castelli, Y. Ioannidis, S. Ross, C. Thanos, P. Pagano, G. Koutrika, H.-J. Schek, y H. Schuldt, "Setting the Foundations of Digital Libraries," *D-Lib*, vol. 13, no. 3/4, Mar. 2007.
- [28] L. Guo, "On construction of digital libraries in universities," *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, 2010, vol. 1, pp. 452–456.
- [29] UIUC, "OAI Registry at UIUC," 2013. [En línea]. Disponible: <http://gita.grainger.uiuc.edu/registry/>. [Última visita: 21-Mar-2013].
- [30] R. Crow, "The Case for Institutional Repositories: A SPARC Position Paper," *The Scholarly Publishing & Academic Resources Coalition*, pp. 1–37, 2002.
- [31] C. A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL*., Feb-2003. [En línea]. Disponible: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. [Última visita: 28-Ene-2013].
- [32] H. Van de Sompel, S. Payette, J. Erickson, C. Lagoze, y S. Warner, "Rethinking Scholarly Communication," *D-Lib Magazine*, vol. 10, no. 9, Sep. 2004.
- [33] J. Tramullas Saz y P. Garrido Picazo, "Software libre para repositorios institucionales: propuestas para un modelo de evaluación de prestaciones," *El Profesional de la Información*, vol. 15, no. 3, pp. 171–181, 2006.
- [34] G. Flores Cuesta y N. Sánchez Tarragó, "Los repositorios institucionales: análisis de la situación internacional y principios generales para Cuba," *ACIMED*, vol. 16, no. 6, pp. 0–0, Dec. 2007.
- [35] A. Bustos González y A. Fernández Porcel, "Directrices para la creación de repositorios institucionales en universidades y organizaciones de educación superior," *Universidad del Rosario*, Ago. 2008.

- [36] J. Xia y D. B. Opperman, "Current Trends in Institutional Repositories for Institutions Offering Master's and Baccalaureate Degrees," *Serials Review*, vol. 36, no. 1, pp. 10–18, Mar. 2010.
- [37] R. Devakos y A. Toth-Waddell, "Ontario Government documents repository D-Space pilot project," *OCLC Systems & Services*, vol. 24, no. 1, pp. 40–47, Feb. 2008.
- [38] CENATIC, "Software de fuentes abiertas en la Administración electrónica Archivo electrónico de las Administraciones Públicas." 2009.
- [39] J.-A. Moreiro, S. Sánchez-Cuadrado, V. Palacios, y E. Barra, "Evaluación de Software Libre Para la Gestión de Archivos Administrativos," *El Profesional de la Información*, vol. 20, no. 2, pp. 206–213, Mar. 2011.
- [40] C. Rodríguez-Vela, "Información de la Administración General del Estado en la era Digital: Publicaciones Oficiales," *El Profesional de la Información*, vol. 20, no. 3, pp. 315–323, May 2011.
- [41] C. Marcos-Martín y S. L. Soriano-Maldonado, "Reutilización de la información del sector público y open data en el contexto español y europeo: Proyecto Aporta," *El profesional de la información*, vol. 20, no. 3, pp. 291–297, 2011.
- [42] R. Gonzalez Castrillo, *Oposiciones a bibliotecas y archivos: escalas de facultativos y de ayudantes : temario básico*. Editorial Complutense, 2002.
- [43] R. Stallman, *Software libre para una sociedad libre*. Traficante de Sueños, 2004.
- [44] P. Suber, "Open access, impact, and demand," *BMJ*, vol. 330, no. 7500, pp. 1097–1098, May 2005.
- [45] P. Suber, "Timeline of the Open Access Movement," 2009. [En línea]. Disponible: <http://www.earlham.edu/~peters/fos/timeline.htm>. [Última visita: 31-Ene-2013].
- [46] OAD, "Timeline - OAD," 2013. [En línea]. Disponible: <http://oad.simmons.edu/oadwiki/Timeline>. [Última visita: 31-Mar-2013].
- [47] DCMI, "Dublin Core Metadata Element Set, Version 1.1," 2012. [En línea]. Disponible: <http://www.dublincore.org/documents/dces/>. [Última visita: 13-Jul-2012].
- [48] Universidad de Cornell, "Clasificación de Metadatos." [En línea]. Disponible: <http://www.library.cornell.edu/preservation/tutorial-spanish/metadata/table5-1.html>. [Última visita: 22-Sep-2012].
- [49] NISO Press, *Understanding Metadata*. 2004.
- [50] CCSDS, *Reference Model for an Open Archival Information System (OAIS):ISO 14721*. 2002.
- [51] UNESCO, "Directrices para la preservación del Patrimonio digital | Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura," 2012. [En línea]. Disponible: <http://www.unesco.org/new/es/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-the-preservation-of-digital-heritage/>. [Última visita: 13-Jul-2012].
- [52] T. Ferreras Fernández, "Preservación digital en repositorios institucionales: GREDOS," 2010. [En línea]. Disponible: <http://gredos.usal.es/jspui/handle/10366/83130>. [Última visita: 13-Jul-2012].
- [53] C. Lupovici, J. Masanès, C. Lupovici, y J. Masanès, *Metadata for the Long Term Preservation of Electronic Publications*. 2000.
- [54] JISC, "Preservation Services for Digital Repositories -PRESERV 2," 2012. [En línea]. Disponible:

- <http://www.jisc.ac.uk/whatwedo/programmes/preservation/preserv2.aspx#>. [Última visita: 13-Jul-2012].
- [55] T. N. Archives, “PRONOM | Welcome,” 2012. [En línea]. Disponible: <http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>. [Última visita: 13-Jul-2012].
- [56] JISC, “KeepIt: Kultur, eCrystals, EdShare (and NECTAR) - Preserve It!,” 2012. [En línea]. Disponible: <http://preservation.eprints.org/keepit/>. [Última visita: 13-Jul-2012].
- [57] PREMIS, “PREMIS: Preservation Metadata Maintenance Activity (Library of Congress),” 2012. [En línea]. Disponible: <http://www.loc.gov/standards/premis/>. [Última visita: 13-Jul-2012].
- [58] B. F. Lavoie, “PREMIS With a Fresh Coat of Paint,” *D-Lib Magazine*, vol. 14, no. 5/6, May 2008.
- [59] PREMIS, *PREMIS Data Dictionary for Preservation Metadata -- version 2.0*. 2008.
- [60] METS, “Metadata Encoding and Transmission Standard (METS) Official Web Site,” 2012. [En línea]. Disponible: <http://www.loc.gov/standards/mets/>. [Última visita: 13-Jul-2012].
- [61] E. Méndez, “Tratamiento de los objetos de información en los archivos: retos y estándares para la descripción basada en metadatos,” 2003. [En línea]. Disponible: <http://eprints.rclis.org/handle/10760/12691>. [Última visita: 13-Jul-2012].
- [62] MODS, “Metadata Object Description Schema: MODS (Library of Congress),” 2012. [En línea]. Disponible: <http://www.loc.gov/standards/mods/>. [Última visita: 13-Jul-2012].
- [63] American Library Association (ALA), *Glosario ALA de Bibliotecología y Ciencias de la Información*. 1983.
- [64] UPEL, *Manual de trabajos de grado de especialización y maestría y tesis doctorales*, 3ra ed. Caracas, Venezuela, 2003.
- [65] M. Barton y M. Waters, “Cómo crear un repositorio institucional. Manual LEADIRS II,” *MIT Libraries*, 2004.

Apéndice A: Informe del Cuestionario

Este cuestionario fue diseñado por el Profesor José Texier, docente adscrito al Departamento de Ingeniería en Informática. Actualmente, cursa un Doctorado en la Universidad Nacional de La Plata (Argentina) cuya investigación está relacionada con las Bibliotecas Digitales y los Repositorios Digitales.

El objetivo de este cuestionario es conocer la gestión y preservación de documentos administrativos (memorandos, oficios, resoluciones, etc.) que se elaboran en las distintas dependencias de la Universidad.

El cuestionario se completó por Internet de forma anónima en el mes de marzo del 2012 y en las siguientes páginas encontrarán un análisis preliminar. Este se envió a 62 jefes de las distintas dependencias administrativas y académicas de la UNET, lo completaron 37, tal como lo muestra la Fig. 1.

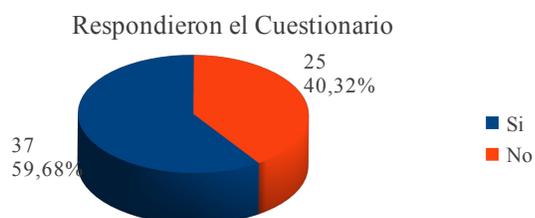


Fig. 1. Completaron el cuestionario

1. ¿En la dependencia que tiene bajo su cargo existe algún proceso de respaldo o copias de seguridad (discos duros externos, memoria USB, correo electrónico, Dropbox, etc.) para los documentos digitales en word, excel, pdf, etc. que tienen?

OPCIONES:

- Sí, en discos duros externos
- Sí, en CD o DVD
- Sí, con copias de seguridad en otros computadores
- Sí, en memorias USB
- Sí, a través del correo electrónico
- Sí, con copias de los documentos y carpetas en el mismo computador
- Sí, con Dropbox o almacenamiento en la nube
- No

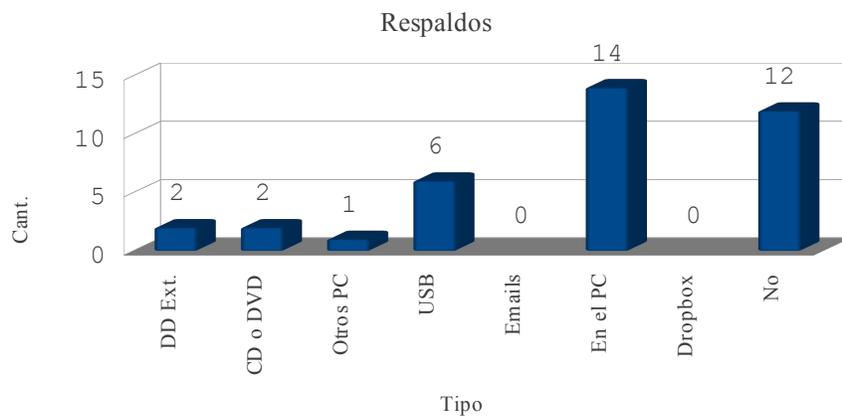


Fig. 2. Detalle de las respuestas de respaldos en fisico.



Fig. 3. Consolidado de respuestas de los respaldos en fisico

En la Fig. 2 se observa la importancia para los responsables de las distintas dependencias académicas y administrativas de la universidad respaldar los documentos (67.57%), eso se comprende porque existen alternativas de almacenamiento en sus dependencias. En la Fig. 3 se realizó un consolidado de las respuestas afirmativas y se comparó con la respuesta negativa.

2. ¿Conoce desde qué año esta dependencia tiene esos documentos administrativos en forma física?

OPCIONES:

- Sí, desde hace un año
- Sí, desde hace 2 años
- Sí, desde hace 5 años
- Sí, desde hace 10 años
- Sí, desde hace 20 años
- No

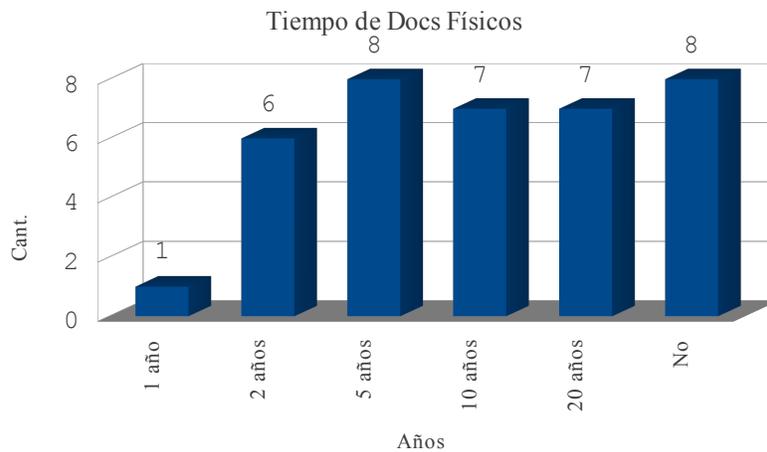


Fig. 4. Detalle de las respuestas de respaldos en digital

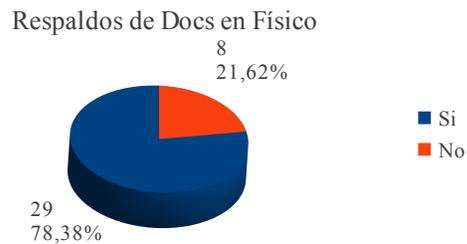


Fig. 5. Consolidado de respuestas de los respaldos en digital

En la Fig. 4 se aprecia que un alto porcentaje archiva los documentos físicos (78.38%) y se observa como las respuestas positivas están distribuidas uniformemente desde la alternativa de 2 años a la de 20 años, esto se infiere por el tiempo en sus respectivos cargos. En la Fig. 5 se realizó un consolidado de las respuestas afirmativas y se comparó con la cantidad de jefes que desconocen desde que año existe un reguardo de sus documentos.

3. ¿Conoce desde que año esta dependencia tiene esos documentos administrativos en forma digital?

OPCIONES:

- Sí, desde hace 1 año
- Sí, desde hace 2 años
- Sí, desde hace 5 años
- Sí, desde hace 10 años
- Sí, desde hace 20 años
- No

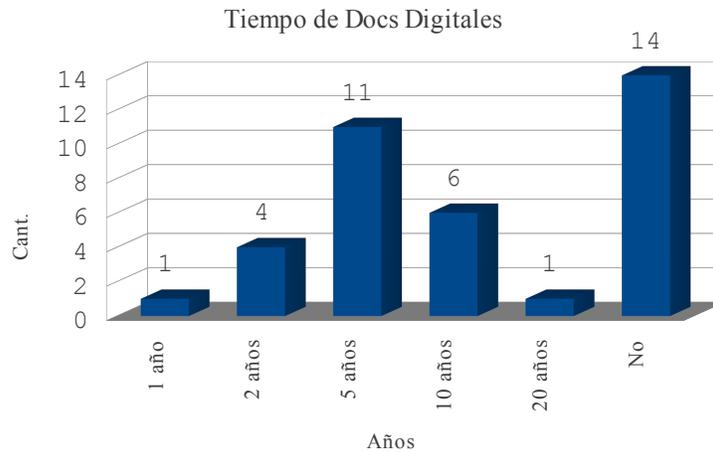


Fig. 6. Detalle por años de documentos en digital

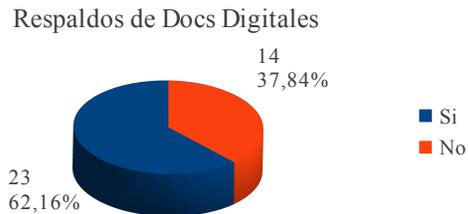


Fig. 7. Consolidado de respuestas de almacenamientos de documentos en digital

En la Fig. 6, al igual que en la Fig. 4, existen muchas dependencias que realizan respaldos de la información, las diferentes respuestas pueden variar por la disponibilidad tecnológica que tenga cada dependencia por diferentes causas. En la Fig. 7 se realizó un consolidado de las respuestas afirmativas y se comparó con la cantidad de jefes que desconocen desde que año existe un reguardo de sus documentos.

4. ¿Todos esos documentos digitales administrativos (en word, excel, pdf, etc) los tienen ordenados o clasificados en la computadora?

OPCIONES:

- Sí
- No

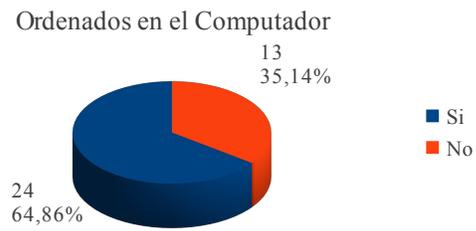


Fig. 8. Relación de documentos ordenados y no ordenados

Nuevamente en la Fig. 8 se aprecia existen muchas dependencias (64.86%) que realizan un ordenamiento o clasificación sobre su información.

5. ¿Conoce algún software que permita recopilar, gestionar, difundir y preservar los documentos administrativos?

OPCIONES:

- EPrints
- DSpace
- Greenstone
- Desarrollos de aplicaciones propias
- Ninguno

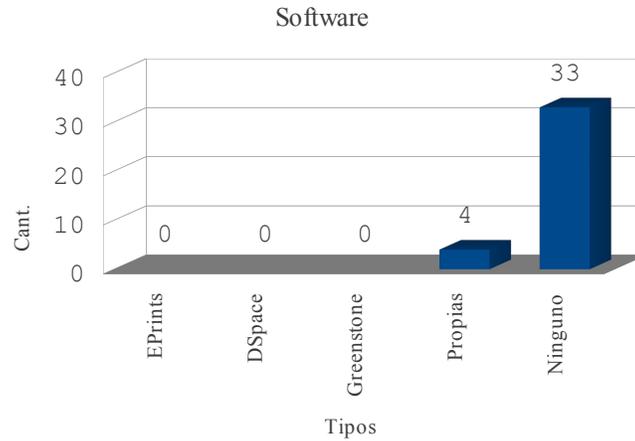


Fig. 9. Detalle por tipos de software conocidos

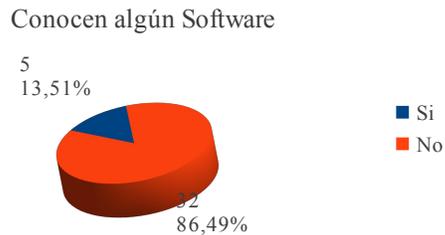


Fig. 10. Consolidado de respuestas por tipos de software

Un porcentaje alto no conoce la existencia de algún software para la administración de los documentos administrativos, según la Fig. 9. A pesar de esto, se puede analizar la importancia que los jefes de las diferentes dependencias de la Universidad le dan a mantener ordenados sus documentos y hacer copias de seguridad por las diferentes formas que ellos conocen. En la Fig. 10 se realizó un consolidado de las respuestas donde se expresa un 86.49% que no conocen plataformas de software para recopilar, gestionar, difundir y preservar esos documentos.

6. ¿Conoce el Decreto de Ley sobre mensajes de datos y firmas electrónicas del 2001 y el Reglamento del 2004, que permite que usted firme de forma digital los documentos administrativos?

OPCIONES:

- Sí
- No



Fig. 11. Relación de conocimiento de la ley

En la Fig. 11 se observa que aproximadamente la mitad de los que completaron el cuestionario conocen la existencia de la ley y del reglamento de mensajes de datos y firmas electrónicas. Esto es muy importante ya que en un futuro todas las dependencias de la Universidad podrían adoptar políticas para firmar electrónicamente los documentos y reducir el uso del papel, además de otros beneficios que se obtienen por la adopción de estos procesos informáticos y electrónicos.

Conclusión

Estos resultados son interesantes para la investigación porque muestran grosso modo cómo y qué se hace con los respaldos de los documentos en las instancias de la universidad, además permitirá realizar una propuesta en los próximos meses para la gestión y preservación de los documentos administrativos, de tal manera, que este más adecuada a la posición de la Universidad y la posición que ustedes tienen desde sus cargos.

Evidentemente se estarán realizando otros cuestionarios y/o encuestas para evaluar propuestas sobre este marco, y poder analizar las fortalezas y debilidades para planificar cambios en los procesos administrativos en nuestra UNET.

Apéndice B: Estructura de la Wiki SeDiCI-DSpace

- Introducción SeDiCI
- SeDiCI
 - Reseña
 - Conceptos sobre repositorios
 - Colaboraciones de SeDiCI con otros repositorios
 - Publicaciones - SeDiCI
- DSpace
 - ¿Por qué DSpace?
 - Características importantes de DSpace
 - Interoperabilidad en DSpace
 - Autoarchivo
 - Ambiente de Desarrollo Recomendado por DSpace
 - Ambiente de Desarrollo DSpace-SeDiCI (SVN:prebi)
 - Aportes de SeDiCI a la plataforma DSpace
 - Ítems del DSpace en SeDiCI
- Configuración Básica DSpace-SeDiCI
 - XMLUI
 - Esquema de Metadatos
 - Comunidades y Colecciones
 - OAI-PMH
 - Discovery
 - Usuarios del Sistema
- Eventos
 - 2012
 - 2011
 - 2010
- Consultas - FAQ
 - @ DSpace
 - @ Eclipse
 - @ Linux
 - Otros links
 - FAQ
- Contactos SeDiCI

Apéndice C: Siglas y Abreviaturas

- API: Application Programming Interface o interfaz de programación de aplicaciones.
- BD: Biblioteca Digital.
- CSS: Cascading Style Sheets o hojas de estilo en cascada
- GD: Gestores Documentales.
- JSPUI: interfaz web de Dspace con soporte la tecnología JavaServer Pages
- MARC: Machine Readable Cataloging o Catalogación Legible por Máquina.
- METS: Metadata Encoding and Transmission Standard.
- MODS: Metadata Object Description Schema.
- OpenDOAR: The Directory of Open Access Repositories.
- OA: Open Access.
- OAI-PMH: Open Archive Initiative Protocol for Metadata Harvesting.
- OAIS: Reference Model for an Open Archival Information System.
- PREMIS: Preservation Metadata: Implementation Strategies.
- RDA: Repositorio de Documentos Administrativos.
- REDAUNET: Repositorio de Documentos Administrativos para la Universidad Nacional Experimental del Táchira.
- RI: Repositorios Institucionales.
- ROAR: Registry of Open Access Repositories
- RSS: Really Simple Syndication.
- SeDiCI: Servicio de Difusión de la Creación Intelectual
- SWORD: Simple Web-service Offering Repository Deposit.
- UNET: Universidad Nacional Experimental del Táchira.
- W3C: World Wide Web Consortium.
- WebUI: Web User Interface.

- XML: eXtensible Markup Language.
- XMLUI: interfaz web de Dspace soportada por Framework Apache Cocoon.
- XSL: Extensible Stylesheet Language o lenguaje extensible de hojas de estilo.