

FACULTAD DE INGENIERÍA

PROYECTO DE GRADO

Repositorios Institucionales

Autores:

Federico ZUBIRI
Andrés MERELES

Supervisor:

Juan José PRADA

24 de mayo de 2012

Resumen

En la actualidad está aumentando la tendencia a universalizar el acceso al conocimiento científico contenido en publicaciones, artículos, y otros tipos de documentos a nivel latinoamericano y mundial. En este sentido, iniciativas como Open Archives Initiative (OAI) proponen políticas y protocolos que son implementados por software (de licencia gratuita o paga) que permiten el acceso abierto a éstas, y lo hacen publicando y compartiendo sus metadatos.

En Uruguay, la Escuela Universitaria de Bibliotecología y Ciencias Afines (EUBCA) está realizando un proyecto que tiene como objetivo implantar políticas de acceso abierto para desarrollo de un repositorio institucional. El proyecto que describe este documento está enmarcado en el proyecto de EUBCA, y tiene como objetivo investigar y/o recomendar un software de licencia libre que cumpla con políticas y protocolos OAI y que se adapte a los requerimientos solicitados por el cliente (equipo de técnicos de EUBCA).

La investigación realizada comenzó explorando la idea detrás de OAI, su forma de implementación y como se lograba interoperar mediante intercambio de metadatos. Además, se investigaron cuales son las tendencias a nivel mundial y latinoamericano en la utilización de políticas OAI. Apoyados por las tendencias investigadas se seleccionaron cuatro software de repositorio institucional (RI) a evaluar.

La evaluación de software candidato se llevó adelante mediante comparaciones cualitativas (funcionalidad), cuantitativas (cantidad de instalaciones, distribución de estas), y en pruebas de instalación, en diferentes plataformas, con diferentes opciones utilizando tecnología de virtualización.

Para asegurar la comparación correcta donde cada aspecto (funcional,

técnico) de cada software fue creada una herramienta que almacena, una calificación numérica de cada aspecto, realiza cálculos y presenta un puntaje final representado en porcentaje. Los resultados obtenidos fueron utilizados como guía para realizar la recomendación del software a utilizar.

Palabras clave:

Repositorio Institucional, interoperabilidad, iniciativas de acceso abierto, protocolos de acceso abierto, metadatos, recolección de metadatos, autoarchivo, software de repositorio.

Contenido

1. Introducción	6
1.1. Definición del problema	6
1.2. Motivación del proyecto	6
1.2.1. ¿Implementación propia o utilización de productos existentes?	7
1.2.2. Características clave de los RIs	7
1.3. Objetivos	7
1.4. Resultados esperados	8
1.5. Cronograma	9
1.6. Organización del documento	11
1.7. Otro material complementario	11
2. Requerimientos del cliente	12
2.1. Requerimientos	12
3. Marco Teórico	14
3.1. Repositorio Institucional	14
3.1.1. Definiciones	15
3.2. Metadatos	16
3.2.1. Dublin Core	17
3.2.2. MARC-XML	17
3.3. Interoperabilidad	18
3.3.1. OAI-PMH	18
Actores clave de OAI-PMH	20
3.4. Mecanismos de depósito	21
3.4.1. Auto depósito (self-archiving)	21
3.4.2. Sistema de depósito centralizado	22
3.4.3. SWORD	22

4. Estado del Arte	23
4.1. Evolución cuantitativa de repositorios	23
4.1.1. Evolución Cronológica	24
Nivel Mundial	24
En Latinoamérica	26
4.1.2. Situación actual	29
Nivel mundial	30
En Latinoamérica	31
4.1.3. Aspectos a tener en cuenta en la elección, puesta en marcha y mantenimiento de un RI	33
Gestión y mantenimiento.	34
Contenidos.	35
Servicios de valor añadido.	36
Visibilidad.	37
Impacto de resultado.	38
4.2. Estudio del sistema Aleph	38
4.2.1. Estudio Inicial	38
4.2.2. Aleph y el protocolo OAI-PMH	42
5. Valoración de Repositorios - Herramienta	43
5.1. Motivación	43
5.2. Descripción	44
5.2.1. Estructura	44
5.2.2. Funcionamiento	46
5.2.3. Elección de los aspectos	50
5.2.4. Conclusiones de los aspectos elegidos y de sus valora- ciones	52
6. Pruebas en Repositorios	54
6.1. Instalación de Repositorios.	54
6.1.1. Guía de instalación EPrints en Ubuntu	54
Requerimientos del producto	54
Preparación del sistema operativo	54
Instalación	55
6.1.2. Guía de instalación de DSpace en Windows XP	55
Requerimientos del Producto	56
Pasos para la Instalación	56
6.1.3. Guía de instalación de DSpace para TEST en EUBCA.	56
Detalles de Instalación	56
Requerimientos del Producto	56
Pasos para la Instalación	57

6.1.4.	Guía de instalación de ZEntity	57
	Detalles de instalación	57
	Requerimientos del producto	57
	Pasos para la instalación	58
6.2.	Estudio de los Repositorios	59
6.2.1.	Estudio sobre DSpace	59
6.2.2.	Pruebas en EPrints	68
7.	Resultados obtenidos	76
7.1.	Evolución del proyecto	76
7.1.1.	Descubrimiento del mundo OAI y los repositorios ins- titucionales	76
7.1.2.	Características comunes de los RIs	77
7.1.3.	Valor agregado en la utilización de la herramienta de valoración de RI	77
7.1.4.	Elección de DSpace como software de repositorio	78
7.2.	Cumplimiento de los requerimientos del Cliente	79
7.2.1.	Interoperabilidad con Aleph	79
7.2.2.	Manejo de metadatos	80
7.2.3.	Control de puntos de acceso	80
7.2.4.	Experiencias en instalación y pruebas de repositorios	81
7.2.5.	Maestro de repositorios: manejo de jerarquías con el protocolo OAI-PMH	82
7.3.	Requerimientos no cumplidos	82
7.3.1.	Proponer mejoras al diseño del RI	82
7.3.2.	Pruebas de carga	83
	Preparación de pruebas de carga	83
8.	Conclusiones	84
8.1.	Trabajos a futuro	85
	Glosario	86
	Referencias	88

Capítulo 1

Introducción

1.1. Definición del problema

En la actualidad la producción científica nacional se encuentra publicada, de manera dispersa en varios sitios, ocasionalmente de acceso restringido y no gratuito, y en otros casos en bibliotecas privadas. Por otro lado, el investigador, estudiante o docente que desee publicar su trabajo no posee un sistema estandarizado, controlado, centralizado, y de acceso abierto a nivel nacional en el cual publicarlo.

Esto ocasiona que las publicaciones sean difíciles de encontrar, dificultando el libre acceso a ellas a cualquier persona que necesite de dicho conocimiento.

Se busca que los investigadores científicos, estudiantes o docentes de nuestro país, que producen material didáctico, científico o tecnológico para todos los niveles del sistema educativo, puedan archivar y distribuir digital y automáticamente su trabajo en una biblioteca centralizada y de acceso público a la cual llamaremos RI.

1.2. Motivación del proyecto

Surge entonces la necesidad de crear una metodología de acceso abierto, la cual pueda ser utilizada como estándar de publicación por los actores que producen el conocimiento, tanto para publicar su trabajo como para que las personas en general puedan acceder a los trabajos publicados.

Solo con la metodología no alcanzaría, sino que en la actualidad, donde se ha masificado el acceso al conocimiento vía Internet, se necesitaría implementar dicha metodología en un software basado en Internet, para así permitir el

acceso abierto.

1.2.1. ¿Implementación propia o utilización de productos existentes?

Es la primera incógnita que este proyecto busca resolver. Para ello se realizará un relevamiento de software de repositorios institucionales (RIs) existentes. Entre las características buscadas estarían las de licencia libre, con buen nivel de soporte y actualización, y con una comunidad de usuarios activa. Se encuentra además como característica esencial, la interoperabilidad con otras tecnologías existentes para publicación de datos. Se investigarán además las tendencias a nivel latinoamericano y mundial de estos productos.

1.2.2. Características clave de los RIs

Se enfocará el estudio de características clave utilizando a la Universidad de la República (UdelaR) como destino de la implementación en calidad de institución de enseñanza e investigación. Es posible que los factores claves varíen si se tratase de otro tipo de institución o empresa con otros fines distintos de la educación.

Los RIs aumentan la visibilidad de los investigadores y de las instituciones: reúnen el producto de la investigación de una institución, lo difunden y lo hacen accesible abiertamente, aumenta la visibilidad del producto, realza la importancia de la investigación que allí se lleva a cabo y aumenta el prestigio de la institución. Aseguran la permanencia en el tiempo del producto y posibilita su utilización como un archivo para la memoria institucional.

Posibilita la gestión de los contenidos digitales de forma más efectiva y transparente. De forma cómoda, rápida y fácil, el investigador puede autoarchivar sus trabajos.

Sirve como apoyo para las tareas de enseñanza y aprendizaje de la institución desde dentro y fuera de la misma, ya que garantiza el acceso abierto desde cualquier otra institución o Internet.

1.3. Objetivos

Nuestro proyecto de grado está inmerso en un proyecto de la EUBCA, el cual tiene como objetivo principal la creación de un repositorio institucional.

Los objetivos del proyecto de EUBCA se separan en generales y específicos.

- *Objetivos generales: Contribuir al desarrollo social, científico y tecnológico del país a través de la promoción del libre acceso a la información y al conocimiento académico producido a nivel nacional.*
- *Objetivos específicos:*
 1. *Determinar cuál es el modelo de comunicación que utilizan los académicos y científicos nacionales para comunicar los resultados de sus investigaciones.*
 2. *Conocer en qué situación se encuentra la producción académica y científica nacional respecto a la libre disponibilidad, facilidad de acceso y situación respecto a los derechos de autor.*
 3. *Analizar las políticas institucionales respecto a la gestión del conocimiento producido en el marco de instituciones financiadas con dinero público.*
 4. *Definir un modelo de repositorio como forma de comunicación científica complementaria que promueva el libre acceso a la información y al conocimiento.*

En este marco los objetivos de nuestro proyecto son:

- ⇒ Realizar un estudio del estado del arte de los RIs disponibles, estudiar/evaluar software de repositorios, con licenciamiento gratuito, los cuales cumplan con requerimientos técnicos y funcionales marcados por estándares existentes y tome en cuenta los requerimientos del cliente.
- ⇒ Evaluar si es viable la utilización de las tecnologías existentes. Recomendar la mejor opción a utilizar o implementar y diseñar una herramienta que permita la evaluación de software de RIs en general.
- ⇒ Diseñar y/o implementar mejoras de alguna funcionalidad o requerimiento del cliente, no existente en el software de repositorio recomendado.

1.4. Resultados esperados

Obtener una visión general de las tecnologías de RIs, evaluar sus características, virtudes y defectos, para poder realizar una recomendación de cual tecnología utilizar en la implementación del RI. Se deberá tener una justificación correcta desde el punto de vista técnico y funcional, y además, cumpla con los requerimientos del cliente.

1.5. Cronograma

A continuación se muestra el cronograma de las actividades que se han realizado en el transcurso del proyecto.

Mes	MAYO				JUNIO				JULIO				AGOSTO			
Semana	1er	2da	3ra	4ta	1er	2da	3ra	4ta	1er	2da	3ra	4ta	1er	2da	3ra	4ta
Investigación estándares																
Investigación productos																
Investigación tendencias																
Herramienta de valoración																
Instalaciones de prueba																
Instalación prototipo																
Reuniones tutoria																
Reuniones con cliente																
Presentación Intermedia																
Presentación final																
Documentación																

Figura 1.1: Cronograma del proyecto, de mayo a agosto del 2011.

Mes	SETIEMBRE				OCTUBRE				NOVIEMBRE				DICIEMBRE			
Semana	1er	2da	3ra	4ta	1er	2da	3ra	4ta	1er	2da	3ra	4ta	1er	2da	3ra	4ta
Investigación estándares																
Investigación productos																
Investigación tendencias																
Herramienta de valoración																
Instalaciones de prueba																
Instalación prototipo																
Reuniones tutoria																
Reuniones con cliente																
Presentación Intermedia																
Presentación final																
Documentación																

Figura 1.2: Cronograma del proyecto, de setiembre a diciembre del 2011.

Mes	FEBRERO				MARZO				ABRIL			
Semana	1er	2da	3ra	4ta	1er	2da	3ra	4ta	1er	2da	3ra	4ta
Investigación estándares												
Investigación productos												
Investigación tendencias												
Herramienta de valoración												
Instalaciones de prueba												
Instalación prototipo												
Reuniones tutoria												
Reuniones con cliente												
Presentación Intermedia												
Presentación final												
Documentación												

Figura 1.3: Cronograma del proyecto, de febrero a abril del 2012.

1.6. Organización del documento

Luego de esta introducción, el documento presente, que constituye el informe, se organiza de la siguiente manera.

En el capítulo 2, se plantean los requerimientos específicos del proyecto. En el capítulo 3, se plantea el marco teórico en el que se basó el resto del proyecto. Luego, en el capítulo 4, se realizó un estudio del estado del arte sobre los repositorios institucionales. En el capítulo 5 se explica la herramienta de evaluación de repositorios institucionales generada por este proyecto. En el capítulo 6, se muestran las pruebas realizadas en los repositorios estudiados. En el capítulo 7, se muestra un resumen de los resultados obtenidos. En el capítulo 8, se plantean las conclusiones del proyecto. Por último, se cuenta con un glosario y referencias bibliográficas utilizadas en el informe.

1.7. Otro material complementario

⇒ Anexo.

Contiene las guías de instalación de cada software candidato y material adicional.

⇒ Herramienta de valoración.

⇒ DVD con instalación de DSpace en Linux Ubuntu.

Capítulo 2

Requerimientos del cliente

En esta sección se muestran los requerimientos relevados para el proyecto.

2.1. Requerimientos

Los requerimientos relevados fueron organizados por tres categorías: Interoperabilidad, Funcional y General. Se listan a continuación según la prioridad que el cliente les adjudicó.

- ⇒ (Interoperabilidad) Integración con el sistema de catálogos de biblioteca de UdelaR llamado Aleph.
 - Estudiar los métodos posibles para poder comunicar el software de RIs con Aleph.
- ⇒ (Funcional) Autogeneración de contenido por el usuario. Sería deseable que el propio usuario que genera, cargue el contenido en el RI sin interacción directa del bibliotecólogo (por lo menos en la parte del ingreso, seguramente deba arbitrar/corregir). Esto está ligado con los siguientes puntos:
 - Definición de metadatos: Explorar alternativas brindadas por los repositorios candidatos.
 - Control de puntos de acceso: ingreso de datos con cierta “calidad”. Lista desplegable los datos correctos. Opciones precargadas.
- ⇒ (Funcional) Relevar complejidad de la infraestructura. Estudio del soporte hardware y/o software, sistema operativo, base de datos, servidor de aplicaciones, servidor web, etc.).

- ⇒ (General) Estudio de tendencias a nivel de América Latina.
- ⇒ (Funcional) Funcionalidad de maestro de repositorios: brindar la capacidad de actuar como centralizador de otros repositorios que eventualmente aparezcan. Proveer/implementar protocolos que lo logren.
- ⇒ (Funcional) Pruebas de carga: estudiar limitaciones de carga para cada repositorio candidato. Proveer alternativas a estas limitaciones.

Capítulo 3

Marco Teórico

En este capítulo se mostrarán definiciones de varios elementos que son utilizados a lo largo del documento.

3.1. Repositorio Institucional

Los RIs surgieron hace aproximadamente diez años como una respuesta de la comunidad académica para manejar, diseminar y preservar sus recursos digitales y han jugado un papel importante en cambios en la publicación y comunicación académica. En particular, los repositorios han estado muy ligados al movimiento de acceso abierto. Sin embargo, no existe una definición única de repositorio institucional, debido en gran parte a divergencias en los objetivos que se pretenden alcanzar con su creación. Esto se ve reflejado en los tipos de materiales académicos que alojan, los mecanismos de depósito que se emplean y los controles de calidad que se establecen.

3.1.1. Definiciones

Para entender el contexto es necesario entender algunos términos y/o metodologías que se nombrarán recurrentemente.

En el caso de los RI, no existe una única forma de definirlos sino que cada definición está condicionada al contexto en el cual es generada.

Presentamos algunas de las definiciones existentes:

Según Lynch (2003) [27]:

Los RIs son una serie de recursos y servicios creados y administrados con el fin de que la comunidad académica pueda manejar y difundir materiales digitales. Su objetivo es capturar y administrar la producción intelectual de una o varias comunidades universitarias y maximizar la visibilidad e impacto en línea.

Según Crow (2002) [28]:

Los RIs son colecciones digitales que capturan la producción intelectual de la comunidad académica y funcionan como indicadores de la producción de la universidad incrementando la visibilidad, status y valor público.

Según Heery et.al (2005) [29]:

Los RIs ofrecen un mecanismo para depositar material por parte del creador, el dueño u otra persona (por ejemplo, un bibliotecario), contar con una arquitectura que maneje tanto el contenido como los metadatos, servicios básicos tales como búsqueda y recuperación, administración, controles de acceso y permisos y ser sustentable a largo plazo, es decir, ser administrado y apoyado por una organización confiable.

Algunos RIs han elegido ampliar la gama de los materiales que reciben para incluir otros tipos de materiales, tales como, tesis, reportes de trabajo, capítulos de libros, memorias de congresos, entre otros. Los tipos de materiales que elige cada repositorio dependen, en gran medida, de la comunidad a la que busca servir, los objetivos de su repositorio y los controles de calidad que quiere establecer. Algunos repositorios utilizan mecanismos externos, tales como la condición de que los materiales deben de estar publicados como indicador de calidad, mientras que otros aceptan cualquier material producido por sus miembros. Hay algunos que aceptan materiales producidos por estudiantes mientras que otros sólo tesis doctorales pero no de licenciatura. Otros han expandido el rango de materiales para apoyar otras áreas del quehacer académico, resguardando documentos administrativos, apuntes de clases, minutas, reportes, entre otros.

Vemos además indispensable definir “iniciativa de acceso abierto”. Definición extracto de “Budapest Open Access Initiative” [14]:

“La literatura que debería accesarse libremente en línea, es aquella que los académicos dan al mundo sin la expectativa de recibir pago. Básicamente, es la categoría compuesta por sus artículos revisados por pares, destinados a publicaciones periódicas; pero también incluye cualquier preprint sin revisión que quizás les gustaría poner en línea para comentar o alertar a otros colegas sobre la importancia de hallazgos de investigación. Hay muchos grados y tipos de acceso amplio y fácil a esta literatura. Por “acceso abierto” a esta literatura queremos decir su disponibilidad gratuita en Internet público, permitiendo a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o usarlos con cualquier propósito legal, sin ninguna barrera financiera, legal o técnica, fuera de las que son inseparables de las que implica acceder a Internet mismo. La única limitación en cuanto a reproducción y distribución y el único rol del copyright en este dominio, deberá ser dar a los autores el control sobre la integridad de sus trabajos y el derecho de ser adecuadamente reconocidos y citados”.

En resumen, los repositorios institucionales deben de contar con mecanismos de depósito, búsqueda y recuperación; ser apoyados por una estructura institucional y contar con políticas de depósito de materiales. Adicionalmente, para lograr interoperabilidad con otros repositorios deben de soportar el protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) para ser recolectados.

La creación de un repositorio no es simplemente un asunto técnico sino una interacción compleja entre diversos grupos de trabajo, tales como informáticos, bibliotecarios, académicos y funcionarios administrativos, quienes en conjunto deben definir políticas y estrategias coherentes y cohesivas que se combinen atinadamente con los intereses de la comunidad a quien interesan servir.

Otros factores tales como: claridad en los objetivos del repositorio, control de calidad, metadatos, asuntos legales, éticos y culturales, culturas de investigación y variaciones entre disciplinas académicas en torno a sus metodologías y prácticas. En consecuencia, los retos principales se encuentran en las diferencias que encontramos en sus objetivos, políticas de ingreso de materiales, controles de calidad y establecimiento de flujo de depósito.

3.2. Metadatos

En las siguientes sub secciones mostraremos dos conjuntos de metadatos utilizados por los softwares de librerías y RIs para poder compartir datos

sobre libros u otros elementos que contengan.

3.2.1. Dublin Core

Dublin Core (DC) es un modelo de metadatos que ha sido creado por Dublin Core Metadata Initiative (DCMI). Las implementaciones de DC usan generalmente XML y se basa en el Resource Description Framework. Se define por: ISO en su norma 15836 de febrero del 2009, estándar ANSI/NISO Z39.85-2007 de mayo del 2007 e IETF RFC 5013 de agosto del 2007.

DC define 15 definiciones semánticas básicas, las cuales pretenden proporcionar información descriptiva básica sobre cualquier recurso. Dichas definiciones son opcionales, pueden repetirse y aparecer en cualquier orden.

Etiqueta	Descripción
Contributor	Entidad responsable de realizar contribuciones al recurso
Coverage	Espacio o tema temporal del recurso, la aplicabilidad espacial del recurso, o la jurisdicción en la cual el recurso es relevante
Creator	La entidad responsable en primera instancia de crear el recurso
Date	Un punto o período de tiempo asociado con un evento en el ciclo de la vida del recurso
Description	Una descripción del recurso
Format	Formato de archivo, medio físico, o dimensiones del recurso
Identifier	Una referencia no ambigua del recurso en un contexto dado
Language	El lenguaje del recurso
Publisher	La entidad responsable de hacer disponible el recurso
Relation	Un recurso relacionado
Rights	Información sobre los derechos del recurso
Source	Un recurso relacionado del cual se deriva el recurso descrito
Subject	El tópico del recurso
Title	El nombre dado al recurso
Type	La naturaleza o género del recurso

3.2.2. MARC-XML

MACHine-Readable Cataloging (MARC) es un formato de datos y un conjunto de estándares asociados, utilizado por las librerías para codificar y compartir información sobre libros y otros materiales. Fue creado en los años sesenta y sigue la especificación ISO 2709 para definir la estructura de cada registro.

MARC21 es el resultado de la combinación de los formatos MARC utilizados en Estados Unidos y Canadá (USMARC y CAN/MARC) y está basado en el estándar ANSI Z39.2, lo que permite la comunicación entre distintas aplicaciones para intercambiar datos.

Se ha desarrollado un esquema XML basado en los estándares MARC21 llamado MARCXML.

3.3. Interoperabilidad

La normalización ha jugado un papel primordial en la creación de repositorios en acceso abierto. Se hacía necesario un estándar para poder intercambiar la información digital contenida en los diferentes archivos abiertos. El objetivo era partir de un mismo esquema de descripción de la información para poder intercambiar los datos. Esto se consiguió con el protocolo OAI-PMH, desarrollado por la Open Archives Initiative y cumplido por las aplicaciones informáticas diseñadas para la puesta en marcha de repositorios y colecciones digitales. Los repositorios que emplean OAI-PMH posibilitan que sus archivos abiertos sean recuperados a partir de recolectores (harvesters), los cuales actúan a modo de meta-buscadores.

Estos meta-buscadores juegan un papel esencial en la visibilidad de presentación/publicación de nuestro conocimiento hacia el mundo. En este sentido es esencial contar con un software que garantice esta funcionalidad.

Algunos de los meta-buscadores más usados son: Google Académico, Google Scholar, HISPANA, DIALNET, RECOLECTA, DRIVER, OAISTER, SCIENTIFIC COMMONS.

3.3.1. OAI-PMH

OAI-PMH es un protocolo desarrollado por Open Archives Initiative. Es usado para recolectar (harvest) descripción de metadatos en un archivo usando metadatos de otros archivos. OAI-PMH utiliza XML sobre HTTP.

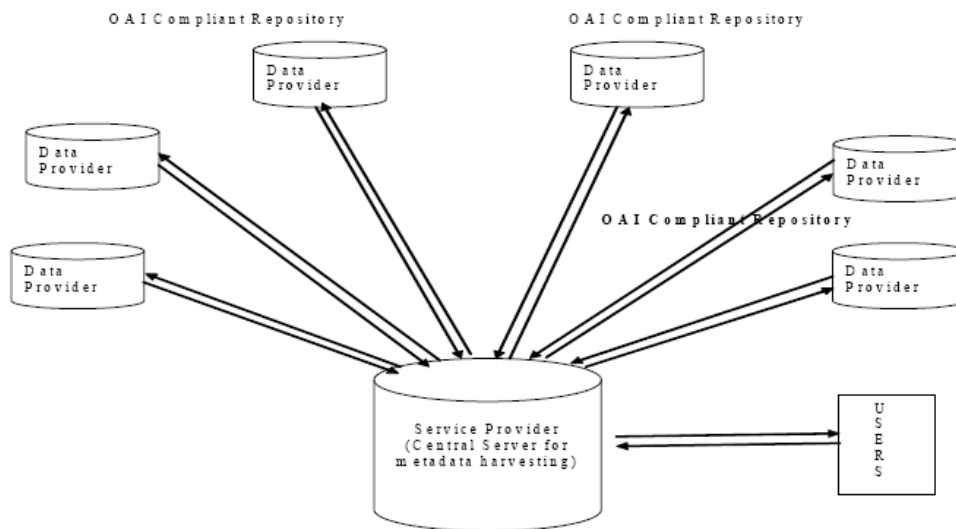


Figura 3.1: Esquema de recolección de metadatos de OAI-PMH.

OAI-PMH ofrece a los proveedores de datos una solución técnica sencilla que posibilita mantener disponible sus metadatos para los servicios que los recolectan, basados en el estándar HTTP y XML. De esta manera los metadatos de muchas fuentes pueden ser reunidos en una base de datos y ofrecidos a los proveedores de servicios para ser consumidos desde ese repositorio centralizado. En la figura 3.1 muestra el esquema de recolección de metadatos.

OAI-PMH es un mecanismo de bajo nivel para interoperabilidad de repositorios. Los proveedores de datos son repositorios que exponen su estructura de metadatos vía OAI-PMH. Entonces los proveedores de servicios hacen llamadas (request) al servicio OAI-PMH para recolectar los metadatos.

OAI-PMH utiliza un conjunto de seis verbos (o servicios) que son invocados desde HTTP mediante operaciones de (*request/replies*). En la figura 3.2 se muestra dicho proceso.

OAI-PMH está basado en los estándares comunes de XML y DC, y también es compatible con MARC. Sin embargo, DC ha sido adoptado como el formato de metadatos común que debería ser soportado por todos los proveedores de datos.

Una de las concepciones erróneas es considerar que la recolección se realiza a nivel de contenido (documentos, ficheros, imágenes). Realmente, OAI-PMH es un protocolo de intercambio de metadatos, por lo tanto, los recolectores incorporan referencias a la localización original de los contenidos (de ahí la importancia de crear/mantener identificadores persistentes). [34]

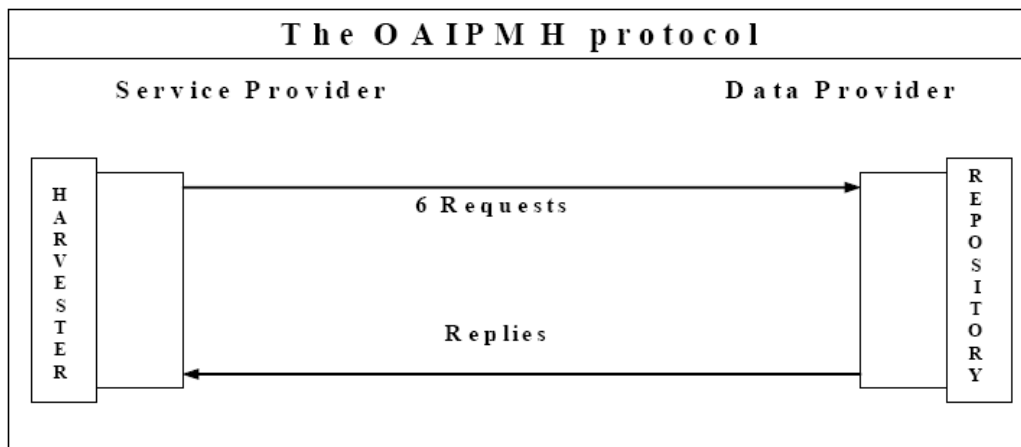


Figura 3.2: Esquema de comunicación (*request/reply*) de OAI-PMH.

Actores clave de OAI-PMH

Se distinguen dos clases de participantes en el protocolo OAI-PMH:

- ⇒ Proveedores de datos (Data providers): mantienen uno o más repositorios (webservers) que soportan OAI-PMH como forma de exponer los metadatos. Estos proveedores procesan y responden las solicitudes recibidas desde los proveedores de servicios (service providers) con mensajes de respuesta OAI-PMH (OAI-PMH response).
- ⇒ Proveedores de servicio (Service providers): Un proveedor de servicios envía solicitudes OAI-PMH (OAI-PMH request) a los proveedores de datos y utiliza los metadatos obtenidos como base para la creación de servicios de valor añadido. Son los recolectores de metadatos que los utilizan con el propósito de ofrecer un servicio de interfaz de búsqueda, sistemas de corrección, etc..

La figura 3.3 muestra la interacción de entre los diferentes tipos de proveedores.

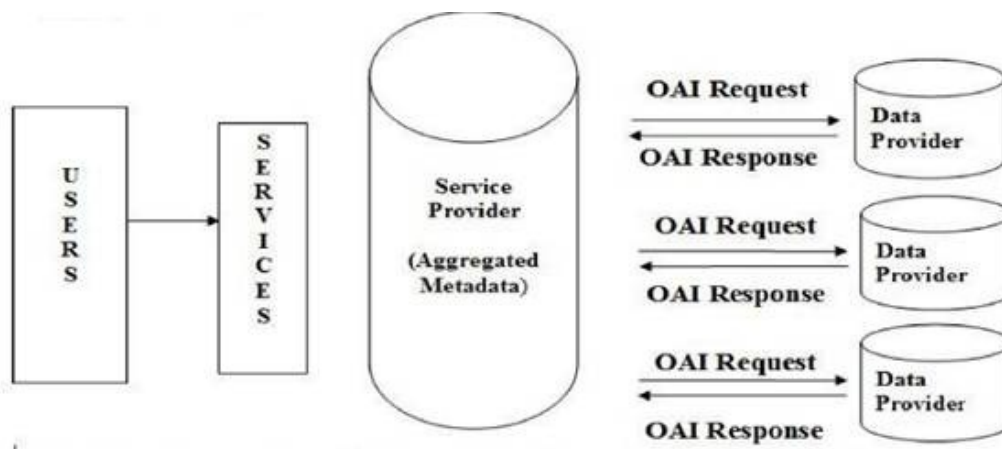


Figura 3.3: Interacción entre proveedores de datos y proveedores de servicios.

3.4. Mecanismos de depósito

3.4.1. Auto depósito (self-archiving)

Este método consiste en que el autor es el responsable de depositar el objeto digital en el repositorio y asignar los metadatos correspondientes. El auto depósito es uno de los caminos posibles para lograr el acceso abierto. Se lo refiere también como depósito sin arbitraje.

Ventajas:

Más efectivo en los repositorios temáticos que para los institucionales, ya que la mayoría reportan poca participación voluntaria. Ésto ha llevado a que en algunas universidades se haya implementado un mandato que obliga a los académicos al auto depósito de sus publicaciones (en particular de artículos) y también mandato de depósito de tesis para los alumnos [1].

Desventajas:

En el auto depósito, no se garantiza la calidad, completitud ni legalidad de lo publicado. El control tanto del contenido y de los metadatos recae en los individuos. Esto podría causar pérdida de calidad en los documentos ingresados. Es importante desarrollar un sistema de acreditación de usuarios autorizados para depositar y otros filtros para lograr que se depositen los materiales deseados de acuerdo con los objetivos del repositorio. Además, se podría implementar un arbitraje a demanda, ante denuncias de contenido inapropiado o erróneo.

3.4.2. Sistema de depósito centralizado

Se lo refiere también como depósito con arbitraje. En este caso una persona o un grupo de personas son los encargados de realizar los depósitos, ingresar los metadatos y asegurar que el tipo de material sea adecuado a las políticas del repositorio. Los materiales depositados pueden ser encontrados de diversas maneras, por ejemplo, revisando la lista de publicaciones de académicos, contactando a los académicos personalmente, recibiendo sugerencias a través de correo electrónico, revisando las páginas personales de los académicos, entre otros.

Ventajas:

Facilidad para implementar controles de calidad para los metadatos y los materiales. Se puede detectar tempranamente metadatos incompletos o duplicación de datos durante la revisión de la publicación.

Desventajas:

Sin embargo, este método requiere de personal o software especializado para realizar las revisiones, lo que eleva los costos de la implementación y manejo de un repositorio. Se posee un mayor tiempo entre el ingreso y la publicación. Adicionalmente se ha criticado que es un sistema poco sostenible a largo plazo, especialmente para instituciones muy grandes.

3.4.3. SWORD

El protocolo Simple Web-service Offering Repository Deposit (SWORD) es usado en repositorios para poder realizar envíos de contenidos desde otras aplicaciones. Es decir, un Servicio Web simple que ofrece servicios de depósito en un repositorio. Nos puede permitir acceder, mediante un servicio web, a realizar envíos directos al repositorio. Lo interesante de esta práctica es que se puede configurar el servicio de envío para poder simplificar u omitir pasos del proceso de envíos y que cualquier usuario registrado o no registrado en el repositorio pueda insertar sus contenidos de forma simple. Posteriormente al depósito, se deberá validar y completar los datos recibidos mediante un mecanismo de arbitraje de contenido.

Como ejemplo de cliente SWORD, se tiene a Authoring Add-in para Microsoft Word. Este cliente permite a los autores y editores guardar archivos en formato Word en la National Library of Medicine's NLM DTD. [37]

Capítulo 4

Estado del Arte

En este capítulo se mostrará el estudio realizado sobre el estado del arte en cuanto a los repositorios institucionales y tecnología asociada a los mismos.

4.1. Evolución cuantitativa de repositorios

En esta sección mostraremos la evolución cuantitativa de los RI, a nivel mundial y Latinoamericano. Se buscará mostrar mediante datos actualizados la actual distribución de RIs, del software que los implementan y las tendencias actuales.

A setiembre de 2011, OpenDOAR [18] recoge un total de 2085 organizaciones con repositorios institucionales.

4.1.1. Evolución Cronológica

Nivel Mundial

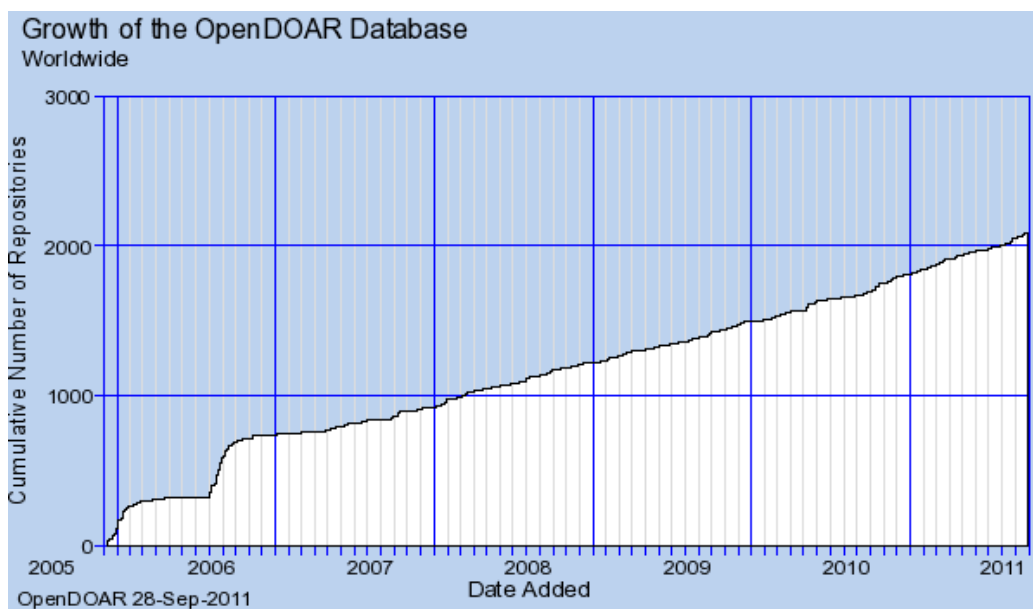


Figura 4.1: N° total de RIs a nivel mundial. Evolución temporal.

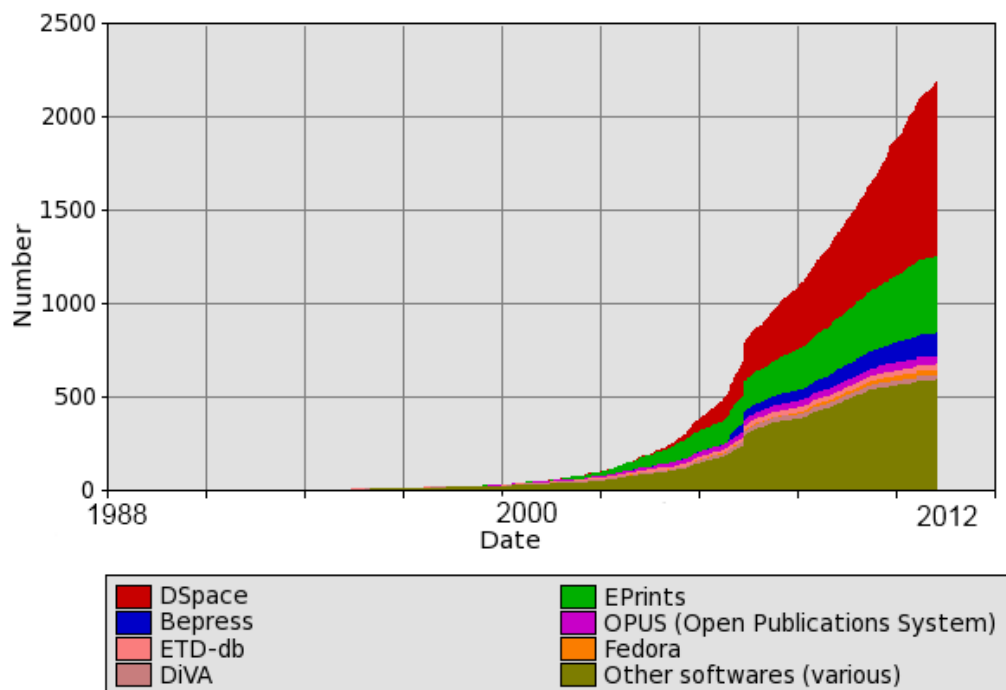


Figura 4.2: Evolución de utilización de software en implementación de RIs a nivel mundial

Tanto en la figura 4.1 como en la figura 4.2 muestran un crecimiento lineal ininterrumpido de la cantidad de repositorios existentes a nivel mundial. Desde el año 2007 a la fecha el crecimiento es de un 25 % anual.

Con respecto al software que implementan los RIs, la figura 4.2 muestra crecimiento desigual de algún software en particular.

Si tomamos el año 2007 como punto inicial de comparación, en ese año había aproximadamente:

- ⇒ 1000 repositorios registrados en openDOAR (fuente de origen de los datos de la figura 2.2) y de estos 1000 RI
- ⇒ 25 % estaban implementados en DSpace,
- ⇒ 25 % en EPrints,
- ⇒ 40 % con otros software (en la figura 4.2 representado por “various”) y
- ⇒ 10 % con el resto de los softwares registrados.

A setiembre de 2011 tenemos:

- ⇒ + de 2000 repositorios registrados en openDOAR, y de estos

- ⇒ 40 % están implementados en DSpace,
- ⇒ 20 % en EPrints,
- ⇒ 30 % con otros software (en la figura 4.2 representado por “various”) y
- ⇒ 10 % con el resto de los softwares registrados.

Se concluye que a nivel mundial:

- ⇒ DSpace es el más utilizado (40%) con 10% de ventaja sobre el 2do grupo (“various”) y un 20% de ventaja sobre EPrints. Teniendo en cuenta que el grupo “various” puede estar formado por gran cantidad de software pero de menor porte.
- ⇒ Se tiene en cuenta a DSpace y EPrints como los preferidos (con 60% de las implementaciones). Se nota que DSpace ha tomado ventaja (duplicando la cantidad de EPrints) y con una tendencia clara a aumentar esa diferencia. Podemos decir entonces que DSpace es el preferido a nivel mundial y que la tendencia es a aumentar esa ventaja.

En Latinoamérica

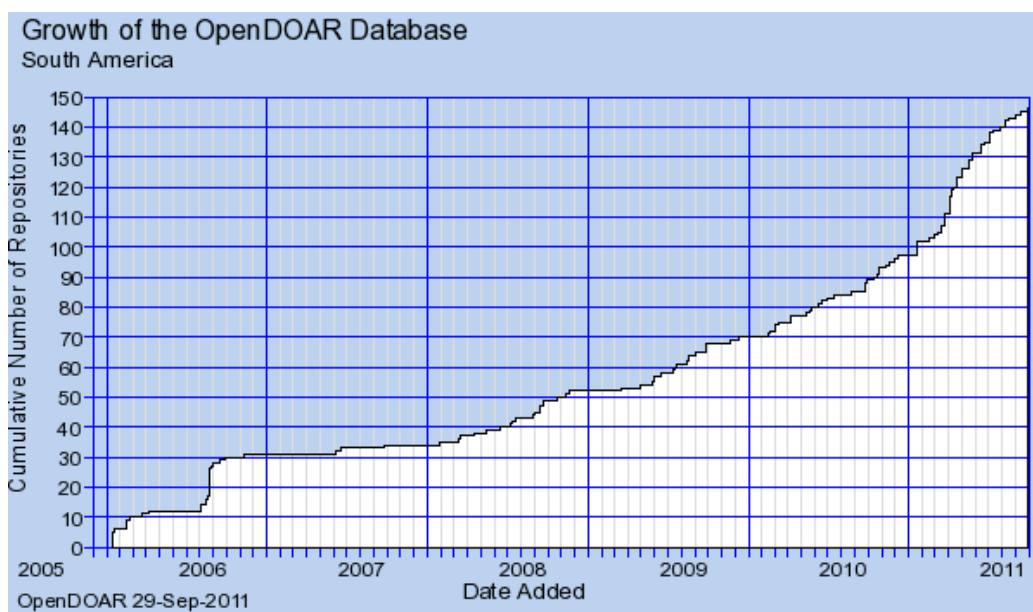


Figura 4.3: RI en Latinoamérica

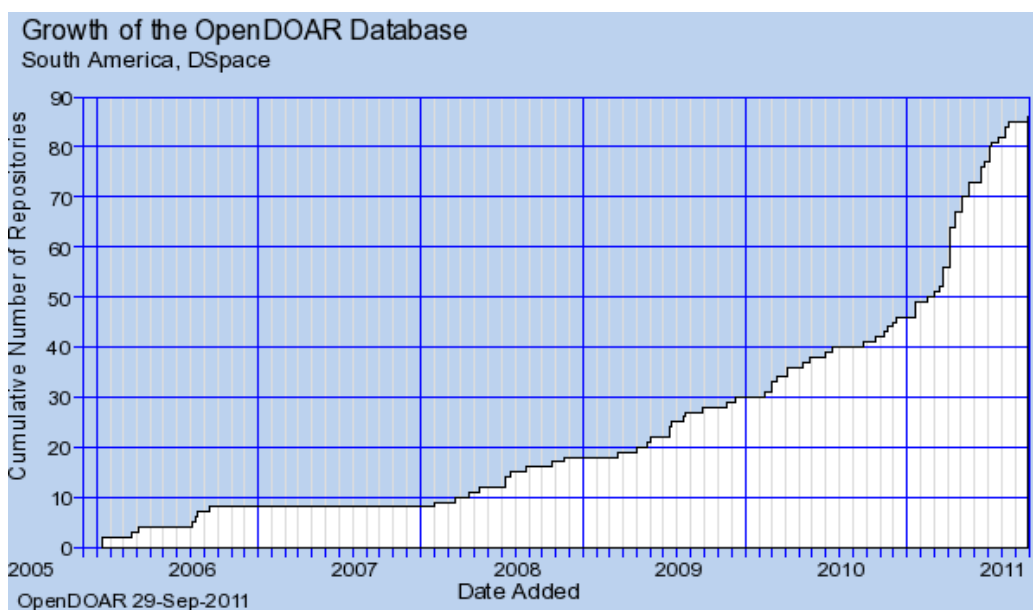


Figura 4.4: RI en Latinoamérica de DSpace

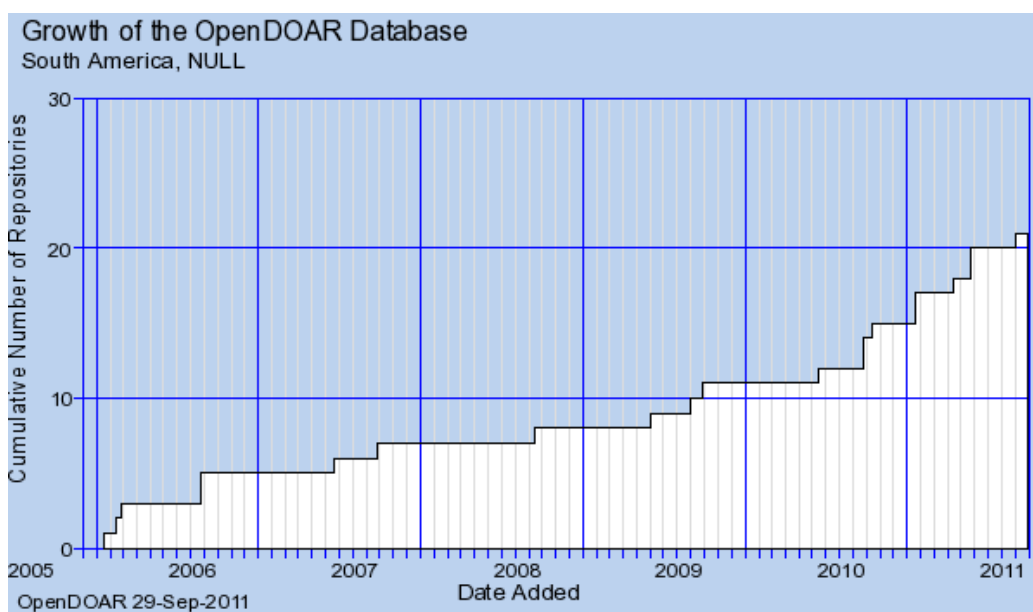


Figura 4.5: RI en Latinoamérica implementados software no identificado

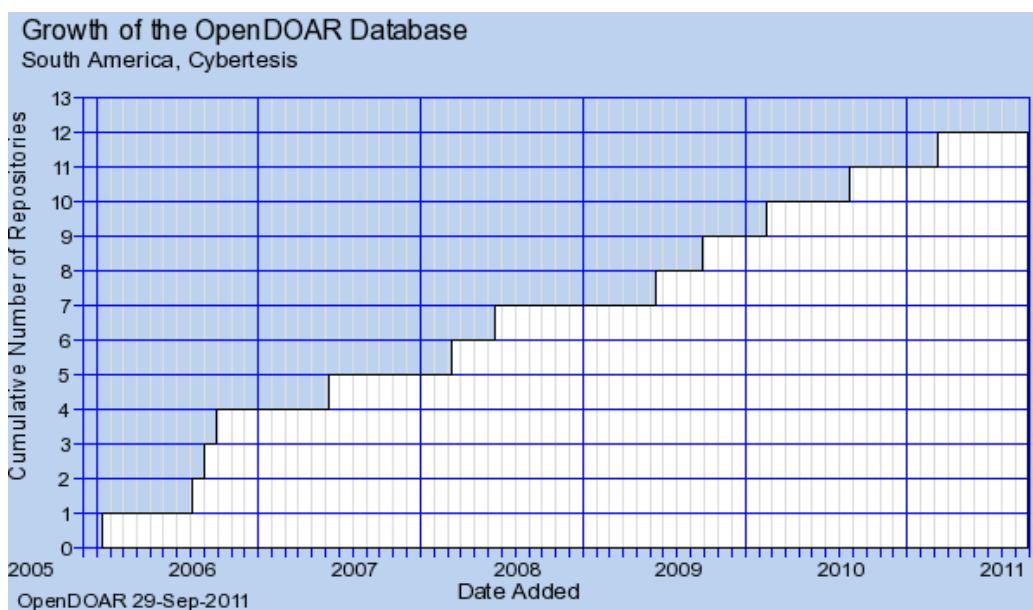


Figura 4.6: RI en Latinoamérica implementados con Cybertesis

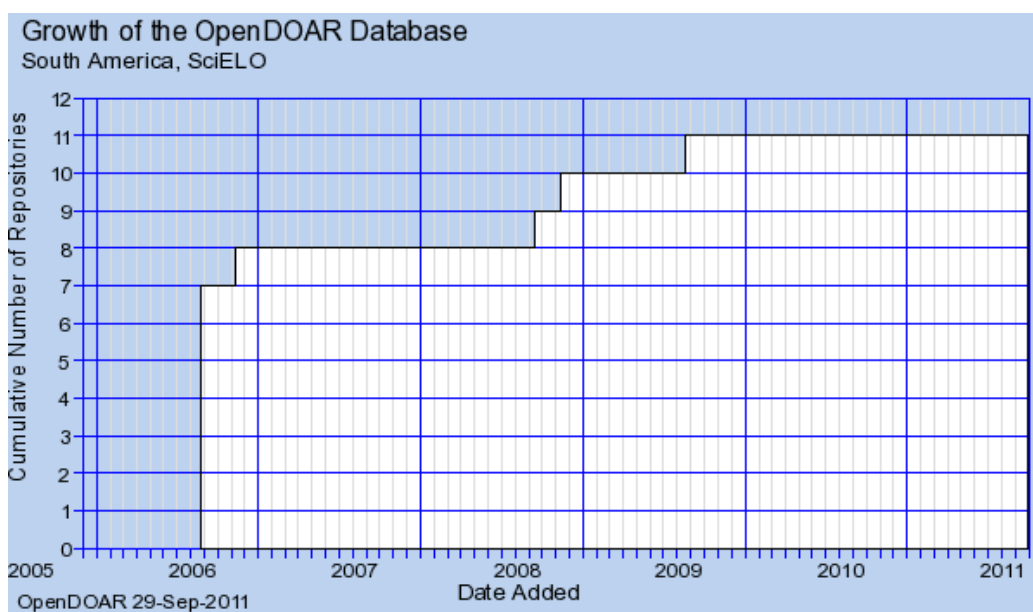


Figura 4.7: RI en Latinoamérica implementados con SciELO

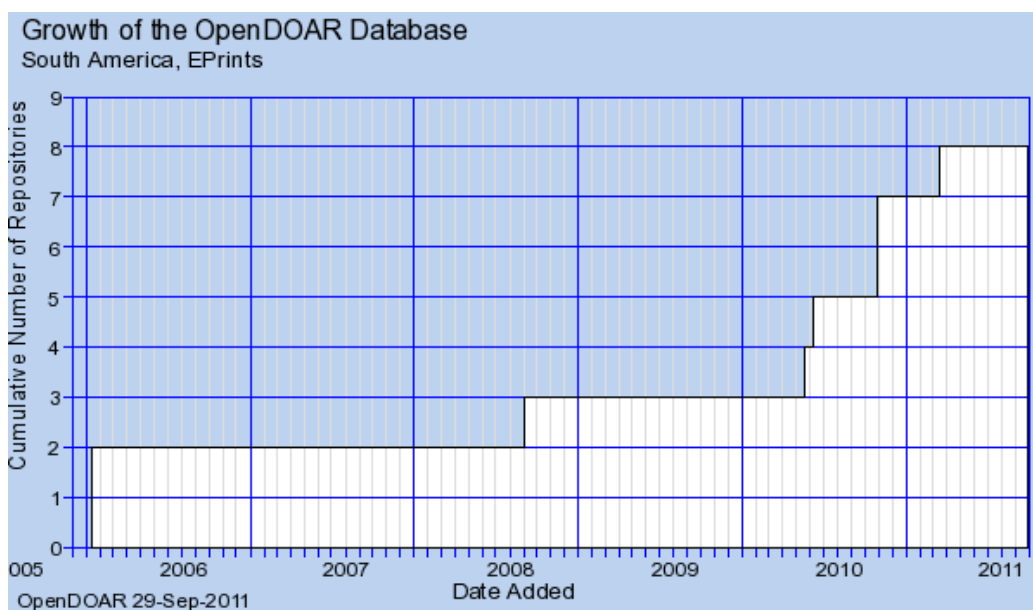


Figura 4.8: RI en Latinoamérica implementados con EPrints

En Latinoamérica, en la figura 4.3, se muestra un crecimiento lineal entre 2008 y 2010 y mayor al lineal en 2011 de la cantidad de repositorios existentes con un crecimiento porcentual mayor al nivel de crecimiento mundial.

Con respecto al software que implementan los RIs, las figuras 4.4, 4.5, 4.6, 4.7, 4.8 muestran el crecimiento de DSpace, “Varios” (software no identificado), Cybertesis, SciELO y EPrints. En estos 5 grupos están representados el 94 % de las instalaciones.

En el caso de Latinoamérica, la supremacía de DSpace es abrumadora, tanto en cantidad de repositorios instalados como en el crecimiento. Actualmente, 86 de los 146 RI instalados utilizan DSpace (59 %). Solo en el año 2011 se instalaron 45 RI utilizando software DSpace contra 8 RI implementados con el resto del software registrado, por lo tanto el crecimiento de instalaciones de DSpace es cercano al 100 % mientras que en el resto de los repositorios fue del 7,5 %.

Por tal motivo DSpace pasó de tener el 45 % de las instalaciones a finales del 2010 a tener el 59 % a la fecha.

Esto marca una tendencia muy clara de cuál es el software preferido.

4.1.2. Situación actual

Obtuvimos de la base de datos de ROAR [19] y openDOAR [18] los datos de cantidad de repositorios correspondientes al mes de setiembre. De estos

datos nos interesa mostrar a continuación dos gráficos comparativos.

Nivel mundial

Proportion of Repositories by Continent
Worldwide

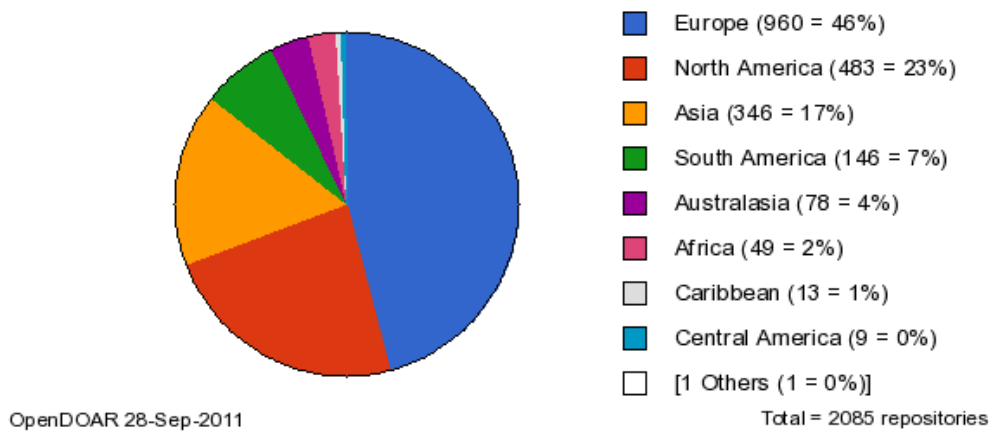


Figura 4.9: N° total de RI por continente.

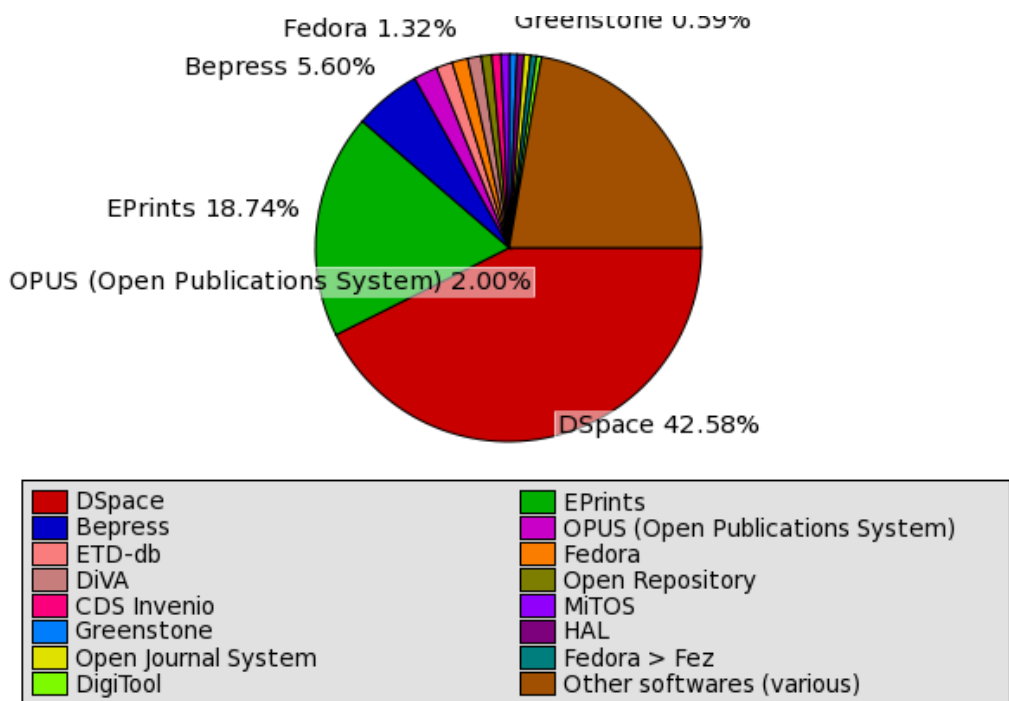


Figura 4.10: % de software utilizado en implementaciones de RI a nivel mundial.

En Latinoamérica

Proportion of Repositories by Country
South America

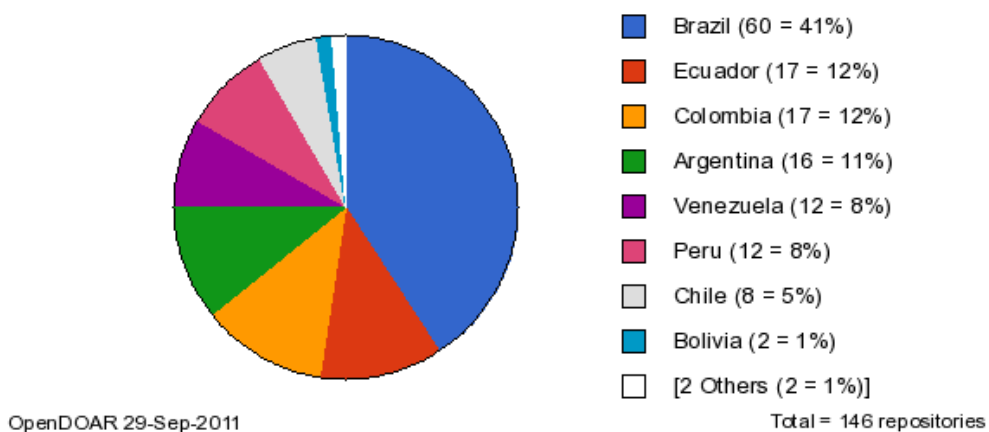


Figura 4.11: N° total de RI en Latinoamérica.

Usage of Open Access Repository Software South America

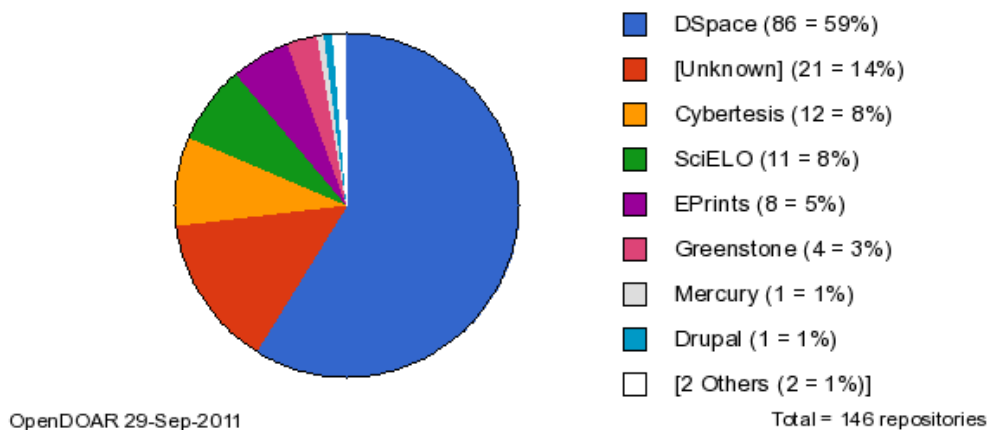


Figura 4.12: % total de software de RI en Latinoamérica.

En la figura 4.9 se muestra la distribución de RI a nivel mundial. Europa y América del Norte cuentan con el 75 % de estas instalaciones y Latinoamérica sólo el 7%. En la figura 4.10 podemos ver la distribución de software que implementa los RI, donde el 60 % de estos utiliza DSpace o EPrints siendo DSpace el más utilizado.

A nivel de Latinoamérica, como se muestra en la figura 4.11, Brasil lleva la delantera con 41 % de los repositorios instalados y la mayoría de las instalaciones también son en DSpace llegando al 59%. A diferencia del resto del mundo, en Latinoamérica, como se muestra en la figura 4.12, el uso de EPrints queda rezagado al 5to lugar con un 5 % y es superado por Cybertesis, SciELO y otros.

Es interesante resaltar que Cybertesis y SciELO solo están presentes en Latinoamérica y por ese motivo influyen tanto en la distribución y no aparecen a nivel mundial.

Proportion of Repository Organisations by Country
Worldwide, Cybertesis

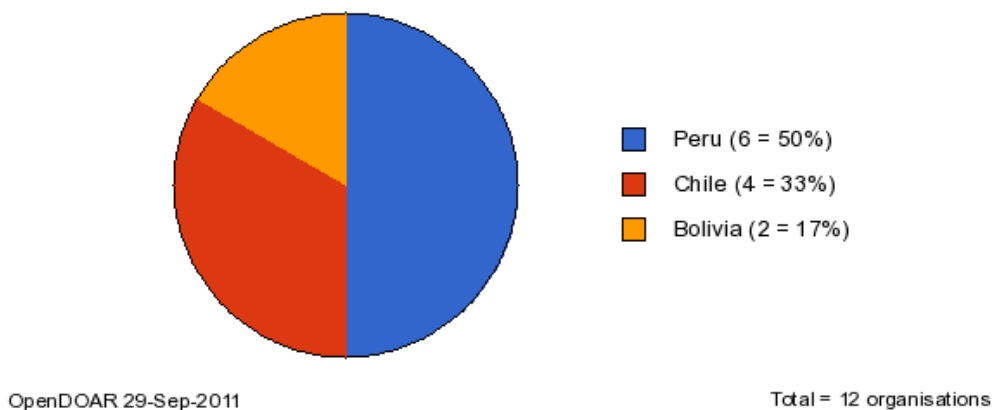


Figura 4.13: Porción de repositorios Cybertesis por país.

Proportion of Repositories by Country
Worldwide, SciELO

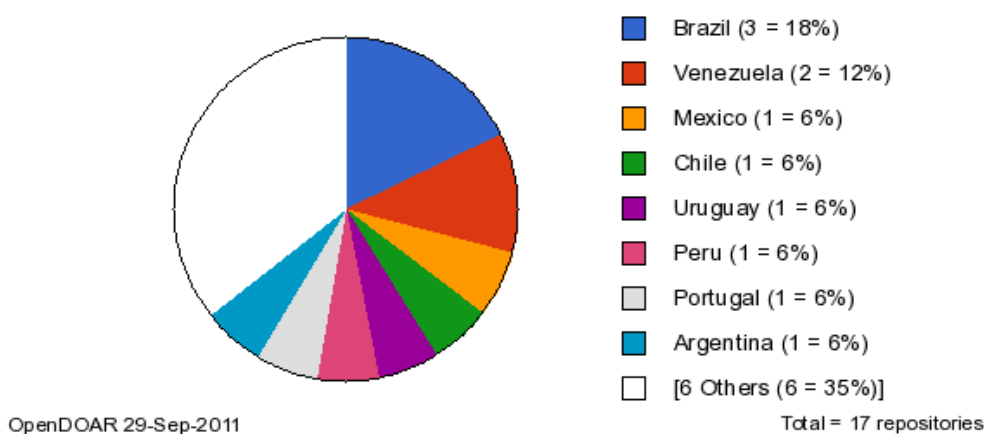


Figura 4.14: Porción de repositorios SciELO por país.

4.1.3. Aspectos a tener en cuenta en la elección, puesta en marcha y mantenimiento de un RI

En esta sección nos interesa mostrar algunos aspectos que pueden influir directamente en la elección de un RI o para estimar la calidad del mismo luego de implementado. Los aspectos que detallaremos a continuación son:

⇒ Gestión y mantenimiento.

- ⇒ Contenidos.
- ⇒ Servicios de valor añadido.
- ⇒ Visibilidad.
- ⇒ Impacto de resultado.

Éstos son complementarios a los factores de evaluaciones Técnicas y Funcionales, las cuales están más relacionados con el software que implementa el RI, y que serán presentados en el capítulo 5.

Gestión y mantenimiento.

En este apartado se enumeraran aspectos correspondientes a la gestión de un RI, que de existir en una organización, aumentan el valor agregado de su utilización.

- Existencia de un manager o administrador.
Favorece a la coordinación de grupos diferentes de trabajo relacionado con el RI. Mantiene una vision general del funcionamiento del RI, sus debilidades y fortalezas.
- Existencia de un equipo de organizado de personas, encargado de gestionar y mantener el servicio.
Favorece la división de tareas y responsabilidades. Genera sinergia. Aprovecha mejor recursos y tiempo.
- Contar con infraestructura consistente y los recursos necesarios para implementar y mantener el servicio.
Indispensable para mantener un nivel de servicio acorde a las necesidades.
- Existencia de parte de la institución de una política clara de apoyo (mandato, recomendación, etc.).
- Existencia de alianzas con otras unidades de la institución (informática, de publicación, académica, etc.) que facilitan la gestión y mantenimiento del servicio.
- Relacionamiento del equipo responsable del servicio con el resto de la plantilla de la biblioteca.
Se considera esencial un buen relacionamiento entre las partes, pues

la gestión correcta de un RI depende de la interacción constante del equipo de soporte y el equipo de bibliotecología.

- Integración del servicio con otros productos de la biblioteca (catálogo, recursos electrónicos, etc.).
En particular, en este proyecto es de especial interés pues se quiere integrar el RI con el catálogo de biblioteca de UdelaR implementado por el sistema Aleph que se describe en la sección 4.2.
- Cursos de formación, tanto para el personal de la biblioteca como para los usuarios.
- Existencia de incentivos para los autores que auto archivan en el repositorio.
Este ítem tiene especial importancia si se quisiera incentivar el auto archivo, por ejemplo, si se está implantando un sistema de RI por primera vez o para sustituir un sistema de almacenamiento existente.
- Costos del servicio residen dentro del presupuesto general de la biblioteca.

Contenidos.

En este apartado se enumeran y comentan aspectos relacionados con el contenido del RI, políticas existentes, restricciones y acceso.

- Contiene diferentes tipologías documentales.
Se refiere a tipos de documentos. Por ejemplo, artículos de autor, monografías (existen por sí mismos, forman una tipología), comentarios de algún artículo de otro autor, catálogo de bibliotecas (existen a partir de otros documentos, forman otra tipología).
- Contiene diferentes formatos.
Refiere a formatos de gráficos como de fuentes.
- Existe relación entre las materias incluidas y las necesidades de los usuarios.
Refiere a la pertinencia de los contenidos con las necesidades del usuario. Puede tener mayor impacto en RIs temáticos que no satisfagan al usuario que los consulta.

- Crecimiento mensual (sostenido) de contenidos.
Se puede tomar como una medida de salud del RI si el crecimiento está dentro de lo esperado, y un motivo de preocupación si no se cumplen los objetivos.
- Incluye restricciones al contenido.
Podrían ser de varios tipos (Por ejemplo, tamaño, temático, formato, etc.).
- Política clara y formal de aportes masivos de documentos o mecanismos para poblar el repositorio.
- Política clara y formal sobre la reutilización o transferencia automática de datos de otros sistemas.
- Política clara y formal sobre el acceso a los contenidos.
Restricciones de acceso, control de acceso.
- Política clara y formal sobre el archivo: que se puede depositar, quien lo puede hacer, en que momentos se puede hacer.
- Política clara y formal sobre los derechos de autor que afectan a los contenidos.
Utilización de protocolos como SHERPA/RoMEO que informan de las restricciones existentes.
- Existencia de mecanismo de control de la calidad de los contenidos (peer review o similar).
- Realización periódicamente de evaluaciones sobre los contenidos.
Definir una política de preservación de contenidos.

Servicios de valor añadido.

En este apartado se enumeran y comentan servicios que salen del estándar en la implementación de RIs pero que pueden mejorar la valoración de un RI como herramienta.

- Incluye módulo de estadísticas de uso (cantidad de citas, producción científica por autores, etc.).
- Servicio de alertas.
Del tipo funcional, que puedan alertar sobre anomalías en el sistema o personalizadas y definidas por los administradores.

- Contiene alguna funcionalidad WEB 2.0 ¹
Incluye gráficos altamente enriquecidos, manejo de streaming ² de audio y video.
- Incluye sistema de clasificación normalizada o vocabulario controlado. Permite control del contenido ingresado, uso de listas negras ³ o grises ⁴ de vocablos.
- Permite el auto archivo.
Utilización del protocolo SWORD para auto archivo en RIs remotos.
- Permite la inclusión de perfiles de los investigadores, con información detallada de cada uno como datos personales, foto, producción científica, especialidad, etc..
- Incluye alguna herramienta de gestión de citas bibliográficas.

Visibilidad.

En esta sección se enumeran y comentan aspectos que influyen en la visibilidad del RI implementado.

- Está presente en directorios nacionales e internacionales.
- Es recolectado por recolectores nacionales e internacionales.
- Mantiene un nombre normalizado que se corresponde con su URL.
- Es indexado por los motores de búsqueda más relevantes.
- Cumple directrices DRIVER⁵.

¹El término Web 2.0 está asociado a aplicaciones web que facilitan el compartir información, la interoperabilidad, el diseño centrado en el usuario y la colaboración en la World Wide Web. Un sitio Web 2.0 permite a los usuarios interactuar y colaborar entre sí como creadores de contenido generado por usuarios en una comunidad virtual, a diferencia de sitios web donde los usuarios se limitan a la observación pasiva de los contenidos que se ha creado para ellos.[30]

²El streaming es la distribución de multimedia a través de una red de computadoras de manera que el usuario consume el producto al mismo tiempo que se descarga.[31]

³Lista de elementos prohibidos.

⁴Lista de elementos desconocidos.

⁵DRIVER, “Digital Repository Infrastructure Vision for European Research” (Visión de infraestructura de repositorios digitales para la investigación europea), es un proyecto realizado por un consorcio financiado por la Unión Europea que está creando un marco de trabajo tecnológico y organizativo para implementar una capa paneuropea de datos, que permita el uso avanzado de los recursos de contenido en el ámbito de la investigación

- Cumple directrices OpenAIRE⁶.

Impacto de resultado.

En esta sección se enumeran aspectos que tienen que ver con los resultados obtenidos luego de la puesta en marcha de un RI y su impacto en el grupo de usuarios y la sociedad.

- Los usuarios están satisfechos con el uso del servicio.
- Es importante la existencia del servicio para la institución.
- Influye la existencia y uso del servicio para que los usuarios alcancen el éxito académico o profesional.
- Es importante este servicio para la sociedad.

4.2. Estudio del sistema Aleph

El sistema Aleph, es un software propietario para el manejo de librerías académicas, de investigación o nacionales construido sobre una base de datos Oracle. Actualmente se utiliza para la catalogación de contenidos bibliográficos de la UdelaR y la Biblioteca Nacional. En este capítulo se hace un estudio de dicho sistema, así como las posibles formas de interacción de dicho sistema con un repositorio institucional.

4.2.1. Estudio Inicial

Se ha propuesto, en una primera instancia, que la obtención de datos del sistema Aleph podría realizarse mediante el post de URL para realizar consultas y un análisis sintáctico (parsing) de página resultado para presentación de datos en alguna interfaz a implementar (del repositorio o a medida).

Se realiza una búsqueda de prueba en el catálogo con la palabra “martí” en categoría “Autor” como se muestra en la figura 4.15:

y la educación superior. DRIVER desarrolla una infraestructura de servicios y una infraestructura de datos. Ambas están concebidas para orquestar los recursos y los servicios existentes en la red de repositorios.[38]

⁶Open Access Infraestructura for Research in Europe, es la disponibilidad inmediata, en línea y gratuita de productos de investigación, sin restricciones de uso comúnmente impuestas por acuerdos de derechos de autor. OpenAIRE incluye los productos que los investigadores liberan normalmente para su libre publicación. Esto incluye, artículos de revistas revisados, documentos de conferencias y conjuntos de datos de varios tipos[39].

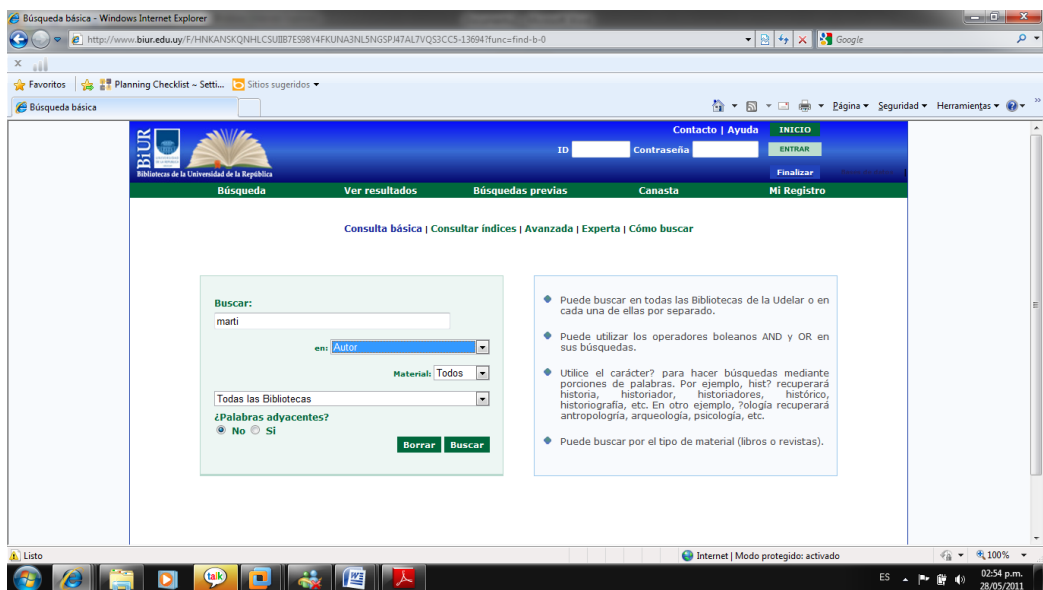


Figura 4.15: Página de Aleph para consultas

Dicha consulta genera el siguiente resultado:



Figura 4.16: Resultado de búsqueda por “martí” por “Autor”

Se puede observar en la figura 4.16 que la URL generada para la búsqueda muestra la palabra buscada y un código que define la búsqueda deseada:

```
http://www.biur.edu.uy/F/HNKANSKQNHLCUIIB7ES98Y4FKUNA3NL5
NGSPJ47AL7VQS3CC5-17127?func=findb&request=marti&find_code=WAU
&filter_code_4=WFM&filter_request_4=&filter_code_5=WSL&
filter_request_5=&adjacent=N
```

Se entiende que modificando los campos `request=marti` y `find_code=WAU` en la URL podría obtener otros resultados. A continuación se muestran las opciones que han sido encontradas en el código fuente HTML de la página de búsqueda:

```
<!-- filename: find-code-include -->
  <select name=find_code>
    <option value=WRD >Todos los campos</option>
    <option value=WTI >Palabras en Títulos</option>
    <option value=WNC >Palabras en Notas de contenido</option>
    <option value=TIT >Título</option>
    <option value=WAU selected >Autor</option>
    <option value=WSU >Materia</option>
  <!-- <option value=UBI >Ubicación</option> -->
    <option value=ISBN >ISBN</option>
    <option value=ISSN >ISSN</option>
    <option value=BAR >Código de barras</option>
    <option value=SYS >Número de sistema</option>
  <!-- <option value=SUD >No. de publicación oficial</option> -->
  <!-- <option value=LCI >LCCN</option> -->
  <!-- <option value=TXT >Texto completo</option> -->
  </select>
```

En la página web resultado de la búsqueda, inspeccionando su código HTML generado, se puede observar como es que son ordenados los distintos elementos contenidos en el resultado de la búsqueda. A continuación se muestra uno de los elementos contenidos en el resultado de la búsqueda en código HTML:

```
<!-- filename: short-a-body-->
<iframe width=0 height=0 id=XX000003 title="XX000003"></iframe>
<tr valign=baseline>
  <td class=td1 id=centered width="1%" valign=top>
    <A HREF=http://www.biur.edu.uy:80/F/HNKANSKQNHLCUIIB7ES98Y4FKUN
    A3NL5NGSPJ47AL7VQS3CC5-09128?func=full-set-set&set_number=100769
```

```

    &set_entry=000003&format=999>3</A>
</td>
<td class=td1 id=centered width="1%" valign=top>
  <form name=XXX000003 method=GET><input type=checkbox name=ckbox
    onClick="javascript:select_record(this.form);"
    onKeyPressed="javascript:select_record(this.form);">
  </form>
</td>
<td class=td1 width="20%" valign=top>Martí, José </td>
<td class=td1 width="5%" valign=top>Monografía </td>
<td class=td1 width="30%" valign=top>Versos sencillos / </td>
<td class=td1 width="5%" valign=top>2009 </td>
<!-- <td class=td1 width="5%" valign=top>861.5&nbsp;MAR ver&nbsp;Cub
</td> -->
  <td class=td1 width="20%" nowrap valign=top>
    <A
      HREF=http://www.biur.edu.uy:80/F/HNKANSKQNHLCUIIB
      7ES98Y4FKUNA3NL5NGSPJ47AL7VQS3CC509129?func=item-
      global&doc_
      library=URE01&doc_number=000186567&year=&volume=&sub_library=
      FHCE
      >Fac.Humanidades y Cs.Educació( 1/ 0)</A><br> </td>
    <td class=td1 nowrap valign=top><br> </td>
    <td class=td1 nowrap valign=top> </td>
  <!-- <td class=td1 nowrap valign=top> </td> -->
</tr>

```

En este fragmento de código HTML del elemento, se puede observar la información referente al mismo, como puede ser el nombre de la publicación o el autor.

Se podría entonces, analizar sintácticamente dicho fragmento de código y presentar la información obtenida. También se podría dar la opción a seguir navegando a través del enlace señalado en rojo, en el cual se muestra mayor información sobre el elemento.

Con esta forma de obtener información del sistema Aleph, la implementación deberá ser revisada en cada nueva actualización hecha en Aleph, para adecuarse a las modificaciones que pudieran existir en su presentación de los datos. Además, existe la posibilidad que en nuevas versiones este método no pueda ser utilizado.

4.2.2. Aleph y el protocolo OAI-PMH

Para la obtención de datos del sistema Aleph se ha estudiado una alternativa con el protocolo OAI-PMH. Aleph soporta peticiones OAI-PMH especificadas en el protocolo OAI. Existe la posibilidad de publicar los datos que contiene un sistema Aleph mediante ALEPH OAI Data Provider. La petición de dicho servicio se realiza mediante una URL y su respuesta es un XML, ambos creada de acuerdo al protocolo OAI. Soporta dos formatos: "MARC21" y "OAI_DC".

Al ser soportado OAI-PMH por los distintos repositorios estudiados, se ha tratado de verificar si el sistema Aleph soporta la utilización de este repositorio, y así poder comunicarse fácilmente entre los sistemas.

Se estableció contacto vía correo electrónico con técnicos de la Servicio Central de Informática (SeCIU), el Servicio Central de Informática, los cuales administran el sistema Aleph de la UdelaR. Su respuesta fue que Aleph en efecto soporta OAI, pero que la UdelaR no ha contratado dicho servicio para su instancia.

Con el protocolo OAI-PMH se pueden obtener resultados de consultas hacia un sistema Aleph y mostrarlas en el sistema repositorio. De esta forma se logra tener la información de catálogo Aleph en conjunto con los datos del repositorio, centralizando las consultas de ambos sistemas logrando así una facilidad de uno por los usuarios mayor. Además, al ser un protocolo estándar soportado por los repositorios estudiados, la comunicación será más estable a los cambios del sistema Aleph y el repositorio a implementar, en sus futuras versiones. Por estas razones se sugiere la obtención del servicio para el sistema Aleph.

Capítulo 5

Valoración de Repositorios - Herramienta

5.1. Motivación

En este punto del proyecto, donde se están evaluando distintos repositorios candidatos y se cuenta con gran cantidad de aspectos a considerar, se entendió indispensable la creación de una herramienta que registre y ordene los aspectos por tipo (aspectos técnicos y funcionales), entendiendo funcionales a los requerimientos que el cliente desea que el repositorio posea (relevados en el capítulo 2). En esta línea, proponemos la creación de una planilla que permita el registro y cálculo de cada aspecto para obtener una valoración numérica y una comparación gráfica de cada aspecto evaluado para cada repositorio candidato.

5.2. Descripción

5.2.1. Estructura

La herramienta propuesta consta de tres espacios de ingreso de datos. Cada uno de los espacios representa un conjunto de aspectos que se agrupan por tipo (técnicos, funcionales y de usuario) y cada aspecto del espacio debe ser valorado con un valor entre 0 y 5, siendo 0 el valor mínimo y 5 el valor máximo.

2		Tecnología disponible/ ambiente de trabajo						
				1				
3		Software Repositorio	Software Soporte		Base de datos	Sub Tot	S F	
4	Dspace		5	5	5	3,7		
5	Eprints			2	5	4		
6	Islandora			1	1	1		
7	Zentity			2	5	4		
8								
9								

Valor Entrada
0 - valor minimo
5 - valor maximo

◀ ◁ ▷ ▶ Asectos_tecnicos Aspectos_funcionales Requerimientos_Usuario Resultados

Figura 5.1: Rango de valores.

Además, cuenta con un espacio donde se presentan resultados calculados a partir de los datos de entrada. Estos resultados son presentados como porcentajes y representan el grado de idoneidad que tiene cada repositorio para cada conjunto de aspectos evaluado.

	%Aspectos_Tecnicos (en el total de tecnicos)	%Aspectos_Funcionales(en el total de funcionales)	%Requerimientos_Usuario
Dspace	70	65	62,59
Eprints	84	75,66	45
Islandora	84	57,61	31,85
Zentity	89	95,73	97,78

	%Tecnicos en el total	%Funcionales en el total	%Total Req_Usuario	%TOTAL
	15,23	23,12	26,71	65
	18,28	26,91	19,2	64
	18,28	20,49	13,59	52
	19,37	34,04	41,73	95
cada categoria --->>	21,76	35,56	42,67	100

Figura 5.2: Resultados en porcentaje.

También se presenta un espacio de gráficas donde se representan los resultados en este formato, con gráficas de barras.

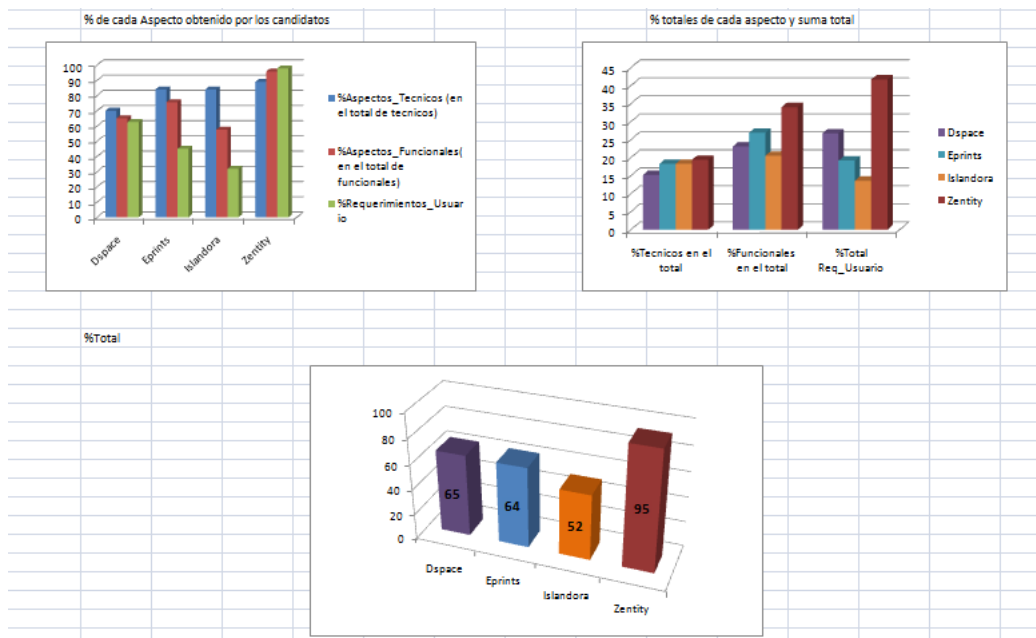


Figura 5.3: Gráficas de barras de los resultados.

En la siguiente sección estudiaremos en detalle cada uno de los espacios y aspectos presentados aquí.

- Aspectos técnicos (Aspectos_técnicos): son los que corresponden a la infraestructura necesaria, licenciamiento, documentación, actualización, seguridad, adaptabilidad del software en sí. Los aspectos fueron seleccionados utilizando guías de evaluación de Software para RI y se agregaron algunos que el equipo de proyecto encontró útil. Entre estos últimos estarían los que están relacionados con la experiencia de instalación, configuración y pruebas que el equipo de proyecto realizó sobre los software candidatos.
- Aspectos funcionales (Aspectos_funcionales): son los que corresponden a la funcionalidad ofrecida por el software candidato. Entre ellos se destacan funcionalidades claves como auto archivo, recuperación de información, información y apoyo en la búsqueda y obtención de información.
- Requerimientos de usuario: son los requerimientos relevados en las reuniones mantenidas con el cliente.
- Resultados: se muestran los resultados (luego de ser procesados) de los datos ingresados en las “hojas” anteriores expresados en porcentajes y agrupados por:
 1. Porcentaje obtenido de los aspectos técnicos.
 2. Porcentaje obtenido de los aspectos funcionales.
 3. Porcentaje obtenido de los requerimientos de usuario.
- Gráficos: representaciones gráficas de los resultados tomando diferentes criterios. Éstos permiten apreciar los resultados obtenidos de una manera más clara.

5.2.2. Funcionamiento

Se ha puesto énfasis en que la herramienta presente resultados acordes a la realidad siendo tan objetivo como se pueda en valorar los diferentes aspectos, pero teniendo en cuenta que muchos de los aspectos están atados a la experiencia, conocimiento previo y/o adquirido durante el estudio, tendencias actuales, realidad actual. Todos estos aspectos aportan subjetivamente a la valoración, lo cual podría ocasionar pérdida de generalidad en el uso de la herramienta pero que aporta para el tipo de proyecto para la cual fue creada.

Se propone evaluar cada aspecto (técnico, funcional, requerimientos) con un valor numérico con 0 como mínimo y 5 como máximo. A su vez cada aspecto está agrupado por tipo y para cada grupo de aspectos se calcula un subtotal (promedio). Ese subtotal será utilizado en el cálculo de resultado.

	Autoarchivo(I)	1							
	Es fácil el registro de usuario para depositar documentos.	Permite modificar el workflow del autoarchivo	Incluye un módulo de administración del autoarchivo	Ofrece soporte del autoarchivo	Incluye la gestión de permisos para el depósito de material.	Permite el autorrelleno.	Ofrece información sobre el depositante (deposited by).	SubTot	
Dspace	5	3	5	0	5	5	5	4	
Eprints	5	5	5	5	2	2	5	4,1429	
Islandora	4	5	5	1	5	2	5	3,8571	
Zentity	5	5	5	5	5	5	5	5	

Figura 5.4: Subtotales.

Se intentó además, hacerla lo más flexible posible, con el objetivo de obtener resultados que representen la realidad que el cliente desea evaluar. Para esto se integró en la herramienta “Factor de multiplicación” y “Multiplicadores de aspectos”:

Factor de multiplicación: Tiene como objetivo modificar los porcentajes máximos alcanzables por cada conjunto de aspectos (Técnicos, Funcionales, Requerimientos de usuario) para que se adapten a alguna realidad que se quiera evaluar. En la concepción inicial de la herramienta este valor se utilizó para hacer que cada uno de los porcentajes máximos alcanzables por cada conjunto de aspectos fuera 33,33% (todos igualados). Para obtener ese valor, los factores utilizados para la hoja Técnicos, Funcionales y Requerimientos de usuario fueron respectivamente 1.18, 3.333334 y 3.333334. Esta desigualdad se daba pues la cantidad de aspectos técnicos que se evalúan casi triplican al resto por lo tanto sin aplicar el factor de multiplicación (idéntico a ponerlos en 1) las factores técnicos ocuparían el 44%, los funcionales un 26% y los requerimientos de usuario un 30%. El rango de valores permitidos

sería entre 1 e infinito. Cabe aclarar que un valor alto le asigna un porcentaje de importancia cada vez más cercano a 100%.

	%Técnicos en el total	%Funcionales en el total	%Total Req_Usuario	%TOTAL		total_tec	total_func	total_req_usu	mult_tec	r
	15,23	23,12	26,71	65		153	250	300	21,7638691	
	18,28	26,91	19,2	64						
	18,28	20,49	13,59	52						
	19,37	34,04	41,73	95						
Maximos en cada categoria --->>	21,76	35,56	42,67	100						

Factor multiplicador			
1,18	3,333334	3,333334	
1,18	3,333334	3,333334	

Figura 5.5: Valoración de subtotales.

Multiplicadores de aspectos: Tiene como objetivo incrementar la importancia de un aspecto en particular. Para esto fue creada una escala entre 1 y 4 y se les otorgó el siguiente significado a cada numero:

1. Valor base.
2. Importante.
3. Muy importante.
4. Esencial.

Internamente lo que la herramienta hace es multiplicar el valor asignado al aspecto por el multiplicador y ese valor es el que cuenta para los resultados.

	Tecnología disponible/ ambiente de trabajo			
			1	
		Importancia del aspecto 1- Valor base (minimo) 2- Importante 3- Muy importante 4- Esencial		
	Software Repositorio	Software Soporte	Base de datos	Su Tc

Figura 5.6: Multiplicadores de Aspectos.

Ejemplo, utilizaremos una parte de la hoja de Aspectos funcionales para mostrar el uso del multiplicador.

Recuperacion (I) priorizados	4								
Ofrece un motor de búsqueda a texto completo.	Permite la búsqueda por metadatos (autor, título, materia, etc.).	Incluye apoyo en la búsqueda (operadores, truncamiento, etc.).	Genera índices.	Incluye protocolos de búsqueda.	Ofrece hipertexto en resultados.	Ofrece la opción de ordenar los resultados (alfabéticamente, cronológico, por relevancia, etc.).	Ofrece diferentes formatos de los registros recuperados (completo, abreviado).	SubTot	
4	5	5	2	1	5	3	1	3,25	
5	5	4	5	5	5	5	1	4,375	
2	5	5	5	1	5	1	5	3,625	
5	5	3	3	5	5	5	5	4,5	

Figura 5.7: Ejemplo de Aspecto

5.2.3. Elección de los aspectos

Continuando con la descripción de la herramienta y considerando que en las secciones anteriores se explicó el funcionamiento de distribución de la misma, nos interesa detallar, y explicar el porqué de la elección de los aspectos y su importancia en la valoración de un RI.

Comenzando con los aspectos técnicos, éstos se obtuvieron de 2 fuentes diferentes: documentación especializada en evaluación de repositorios y experiencia del equipo de proyecto. Se detallan los grupos de aspectos obtenidos de la documentación especializada, su composición y lo relevante a la hora de la valoración:

- *Instalación*, incluye facilidad de instalación y existencia de documentación de apoyo.
- *Solidez, Documentación*, que incluye existencia de una comunidad de desarrollo y activa de usuarios, documentación técnica y de uso, y la existencia de canales de comunicación con las comunidades antes nombradas.
- *Características*, en las que se detallan, autonomía, separabilidad, transparencia, flexibilidad.
- *Permite infraestructura virtual*. Tanto para el sistema en sí y el repositorio.
- *Permite diferentes lenguajes*. Tanto para el sistema en sí y el repositorio.
- *Facilidad de migración*. Tanto para el sistema en sí y el repositorio.
- *Protocolos y tecnologías soportadas para interoperabilidad*, incluye protocolos OIA-PMH, OAIS, SWORD, SOAP. Este grupo de aspectos es de especial importancia para la valoración pues los aspectos que contiene son sensibles a la elección del software a utilizar.
- *Permite extender las funcionalidades mediante plugins y scripts*. Tanto para el sistema en sí y el repositorio.
- *Seguridad*, que incluye validación de acceso y utilización de protocolos LDAP y SSO.
- *Interfaz, acceso*, que incluye la utilización de estándares de metadatos (DC, METS, MODS, MARC21), exportación e importación de datos en masa, identificación persistente de documentos, gestión de perfiles

de usuario y roles, posibilidad de ingreso en diferentes formatos de metadatos y plantillas. Este grupo de aspectos es de especial importancia para la valoración pues los aspectos que contiene son sensibles a la elección del software a utilizar, además se destaca la gestión de metadatos como aspecto fundamental requerido por el cliente.

A continuación, se detallan los grupos de aspectos sugeridos por el equipo de proyecto, su composición y lo relevante a la hora de la valoración:

- *Tecnología disponible - ambiente de trabajo*, que incluye el software de repositorio, el software de soporte y la base de datos.
- *Frecuencia de actualización*, que incluye el software de repositorio, el software de soporte y la base de datos.
- *Licenciamiento*, que incluye el software de repositorio, el software de soporte y la base de datos.
- *Configuración*, que incluye, conocimientos necesario, dificultades de configuración y dificultades de adaptación.
- *Desempeño*, que incluye, tolerancia a carga de usuarios, tolerancia a carga de transacciones y utilización de storage.
- *El hardware requerido ajusta con el hardware disponible?*, que incluye el software de repositorio, el software de soporte y la base de datos.

Continuando con los aspectos funcionales, éstos se obtuvieron de documentación especializada en evaluación de repositorios. Se detallan los grupos de aspectos obtenidos de la documentación especializada, su composición y lo relevante a la hora de la valoración:

- *Autoarchivo*, que incluye, facilidad de registro de usuarios, modificación del workflow de acceso, incluye modulo de gestión de autoarchivo, incluye gestión de permisos y autorrelenado. También se incluye la gestión de derechos de autor como aspecto esencial.
- *Recuperación*, que incluye, apoyo en la búsqueda (utilización de operadores), motor de búsqueda a texto completo y búsqueda avanzada, permite búsqueda por metadatos, genera índices, incluye protocolos de búsqueda, hipertexto en resultados, ofrece diferentes formatos en los registros recuperados, admite browsing, recupera número de registros.

- *Información-Apoyo*, que incluye, información sobre políticas editoriales (ROMEO, DULCINEA), incluye información sobre el servicio, posee tutoriales de autoarchivo y derechos de autor, posee buzón de sugerencia o mail.

Finalmente los requerimientos de usuario. Éstos se obtuvieron de reuniones con el cliente. Se detallan los grupos de aspectos obtenidos, su composición y lo relevante a la hora de la valoración:

- *Interoperabilidad*, en este caso se considero a interoperabilidad en general que incluye la utilización de protocolos como OAI-PMH o similares y la interoperabilidad particular con el sistema ALEP en el actual estado de implantación en UdelaR.
- *Capacidad de funcionar como maestro de repositorios*, o sea, administrar repositorios desde otros repositorios, harvesting a otros repositorios.
- *Autogeneración de contenido*, que incluye los diferentes ciclos de generación de documentos, con validación y sin validación.
- *Control de puntos de acceso*, que incluye en control de usuarios y control de ingreso de datos (no permitir datos erróneos o duplicados).
- *Tendencias a nivel de latinoamérica*, evolución cuantitativa, funcionalmente en línea con lo que se está haciendo en latinoamérica.

5.2.4. Conclusiones de los aspectos elegidos y de sus valoraciones

Hasta aquí se ha detallado el funcionamiento de la herramienta, su estructura y las diferentes formas de configurarla modificando los valores de los *multiplicadores de aspecto* y del *factor de multiplicación*.

Nos resta comentar como se ha configurado la herramienta para la evaluación de software de repositorio para este proyecto. Inicialmente fueron ingresados todo los aspectos a considerar en su correspondiente grupo. Para cada uno de estos grupos se les asignó 1 en el campo *multiplicador de aspecto*, por lo tanto, como los la cantidad de aspectos técnicos supera a los funcionales y requerimientos de usuario, por lo tanto tiene mayor ingerencia en el resultado final. Expresado en numeros y en esta situación, los aspectos técnicos podrían sumar 58.62 puntos sobre 100 y los funcionales y requerimientos solo el el 20.69 sobre 100.

Como la situación anterior podría no representar la realidad que se quiere evaluar, se realizó la corrección de estos valores adecuando el *factor de multiplicación* hasta lograr que cada aspectos técnicos, funcionales y requerimientos de usuario puedan obtener 33,33 puntos sobre 100.

Entonces, lo siguiente fue adecuar los *multiplicadores de aspecto* (inicialmente valiendo 1) para que reflejen la importancia que cada grupo de aspecto tiene. Esta adecuación se llevó a cabo con la ayuda del grupo de proyecto de EUC-BA que identificaron los aspectos a priorizar. El resultado obtenido fue, los aspectos técnicos podrían sumar 23,08 puntos sobre 100, los funcionales 34.97 sobre 100 y requerimientos 41.96 sobre 100. De este escenario se desprende que el cumplimiento de los requerimientos de usuario es prioritario seguido de los aspectos funcionales.

Finalmente, se completó cada uno de los aspectos con un valor entre 0 y 5 tomando como referencia 0 cuando no cumple o no aplique y 5 cuando cumple completamente con el aspecto evaluado. La valoración de cada aspecto se basó en la experiencia obtenida durante la ejecución del proyecto (principalmente los técnicos y alguno de los requerimientos de usuario) y en la documentación, experiencia de usuarios, guías de uso, etc, cuando no se pudo hacer una comprobación empírica del aspecto (principalmente utilizada para los aspectos funcionales).

Capítulo 6

Pruebas en Repositorios

En este capítulo se detallan las tareas de instalación y pruebas que se realizaron sobre el software candidato de RI.

6.1. Instalación de Repositorios.

En esta sección mostraremos el resumen del proceso de instalación de cada producto. Omitimos presentar aquí los detalles técnicos de cada instalación. Estos se pueden ver en el Anexo del informe.

6.1.1. Guía de instalación EPrints en Ubuntu

Se utilizó para la instalación EPrints 3.2 en Ubuntu 11.04 desktop.

Requerimientos del producto

- ⇒ Apache 2.0 o posterior.
- ⇒ MySQL 5 o posterior.
- ⇒ ModPerl 2.0 o posterior.

Preparación del sistema operativo

En este caso se instaló el SO en una máquina virtual VMware Workstation.

Instalación

1. Instalar el software detallado en “*Requerimientos del Producto*”.
2. Obtener fuentes de EPrints desde el sitio web.
3. Instalar fuentes (mediante apt-get de Ubuntu).
4. Configuración de repositorio mediante comandos de EPrints (epadmin create).

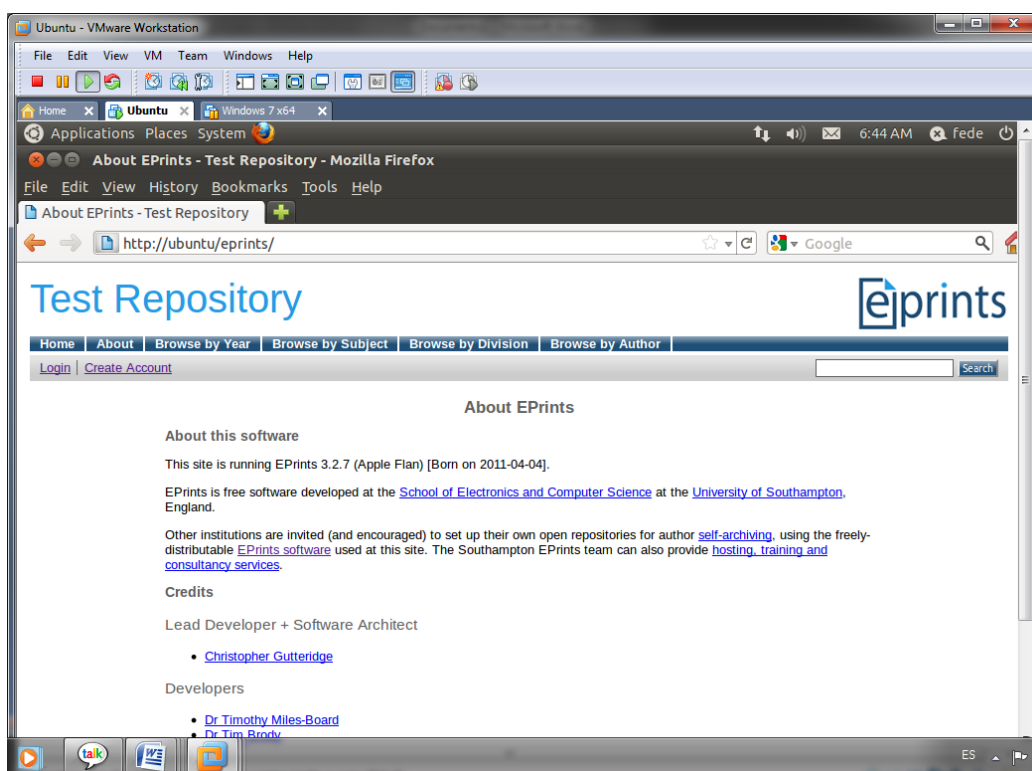


Figura 6.1: EPrints instalado.

6.1.2. Guía de instalación de DSpace en Windows XP

Instalación realizada en el Sistema Operativo Microsoft Windows XP [5.1.2600].

Versión de DSpace utilizada: 1.7.2.

Requerimientos del Producto

- ⇒ Oracle Java JDK 6 o superior.
- ⇒ Apache Maven 2.2.x.
- ⇒ Apache Ant 1.7 o superior.
- ⇒ PostgreSQL 8.2 a 8.4.
- ⇒ Apache Tomcat 5.5 o 6.

Pasos para la Instalación

1. Instalar y configurar software detallado en “*Requerimientos del Producto*”.
2. Bajar del sitio el de DSpace el instalador ejecutable del mismo.
3. Instalar DSpace.
4. Levantar servicios de publicación (Tomcat).

6.1.3. Guía de instalación de DSpace para TEST en EUBCA.

Detalles de Instalación

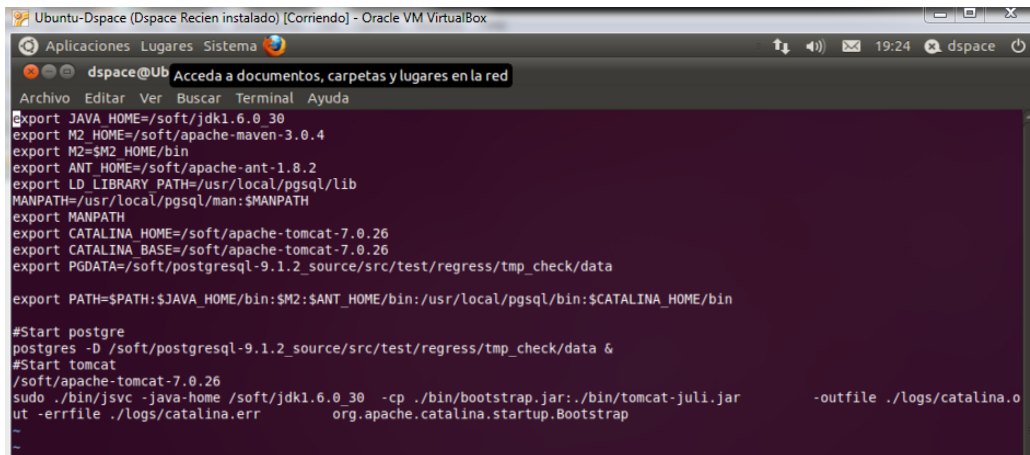
Instalación realizada en el Sistema Operativo Linux Ubuntu 11.04 sobre un virtualizador VirtualBox de Oracle.
Versión de DSpace utilizada: 1.8.2 (Último release estable a febrero de 2012).

Requerimientos del Producto

- ⇒ Oracle Java JDK 6 o superior.
- ⇒ Apache Maven 2.2.x o superior.
- ⇒ Apache Ant 1.7 o superior.
- ⇒ PostgreSQL 8.2 o superior.
- ⇒ Apache Tomcat 5.5 o superior.

Pasos para la Instalación

1. Instalar y configurar software detallado en “*Requerimientos del Producto*” según manual del producto.
2. Bajar del sitio el de DSpace los fuentes a instalar.
3. A diferencia de la instalación en Windows, en Linux se deben compilar los fuentes con gcc (provisto por Ubuntu).
4. Generación de paquete de instalación y ejecución del mismo.
5. Carga de variable de entorno, ejecución de PostgreSQL y Tomcat, mediante ejecución del script llamado *.profileDspace* que se muestra en la figura 6.2.



```
Ubuntu-Dspace (Dspace Recien instalado) [Corriendo] - Oracle VM VirtualBox
Aplicaciones Lugares Sistema
dspace@Ub Acceda a documentos, carpetas y lugares en la red
Archivo Editar Ver Buscar Terminal Ayuda
export JAVA_HOME=/soft/jdk1.6.0_30
export M2_HOME=/soft/apache-maven-3.0.4
export M2=$M2_HOME/bin
export ANT_HOME=/soft/apache-ant-1.8.2
export LD_LIBRARY_PATH=/usr/local/pgsql/lib
MANPATH=/usr/local/pgsql/man:$MANPATH
export MANPATH
export CATALINA_HOME=/soft/apache-tomcat-7.0.26
export CATALINA_BASE=/soft/apache-tomcat-7.0.26
export PGDATA=/soft/postgresql-9.1.2_source/src/test/regress/tmp_check/data
export PATH=$PATH:$JAVA_HOME/bin:$M2:$ANT_HOME/bin:/usr/local/pgsql/bin:$CATALINA_HOME/bin
#Start postgres
postgres -D /soft/postgresql-9.1.2_source/src/test/regress/tmp_check/data &
#Start tomcat
/soft/apache-tomcat-7.0.26
sudo ./bin/jsvc -java-home /soft/jdk1.6.0_30 -cp ./bin/bootstrap.jar:./bin/tomcat-juli.jar -outfile ./logs/catalina.o
ut -errfile ./logs/catalina.err org.apache.catalina.startup.Bootstrap
--
```

Figura 6.2: Script de inicio.

6.1.4. Guía de instalación de ZEntity

Detalles de instalación

Instalación realizada Sistema Operativo Microsoft Windows 7 (x64) [6.1.7601].

Versión de ZEntity utilizada: 2.0.

Requerimientos del producto

⇒ Sistemas operativos soportados:

- Microsoft Windows Server 2008 (x86 y x64) o superior.
- Microsoft Windows 7 Ultimate/Enterprise (x86 o x64).

Se utilizó Microsoft Windows 7 x64.

- ⇒ Microsoft .Net Framework 4.0 Full Version.
- ⇒ Internet Information Server 7 o superior. Se utilizó Internet Information Server 7.
- ⇒ Silverlight 4.
- ⇒ Windows SDK for Windows 7.
- ⇒ Microsoft SQL Server 2008 Native Client.
- ⇒ Windows PowerShell 2.0 Software Development Kit (SDK).
- ⇒ Habilitar Full-Text Search.

Pasos para la instalación

1. Instalar y configurar software detallado en “*Requerimientos del Producto*” según manual del producto.
2. Descargar el instalador del server “**Zentity.Installer.Server.msi**” desde el sitio de Zentity.
3. Ejecutar el instalador y seguir los pasos indicados en el manual.
4. Descargar el instalador del cliente “**Zentity.Installer.Client.msi**” desde el sitio de Zentity.
5. Ejecutar el instalador y seguir los pasos indicados en el manual.

6.2. Estudio de los Repositorios

En la siguiente sección se mostraran los estudios realizados sobre EPrints y DSpace. Estos RI fueron elegidos sobre los cuatro softwares de RI candidatos pues son los que dominan a nivel mundial y latinoamericano, basado en el estudio cuantitativo de software realizado en el capítulo 4.

6.2.1. Estudio sobre DSpace

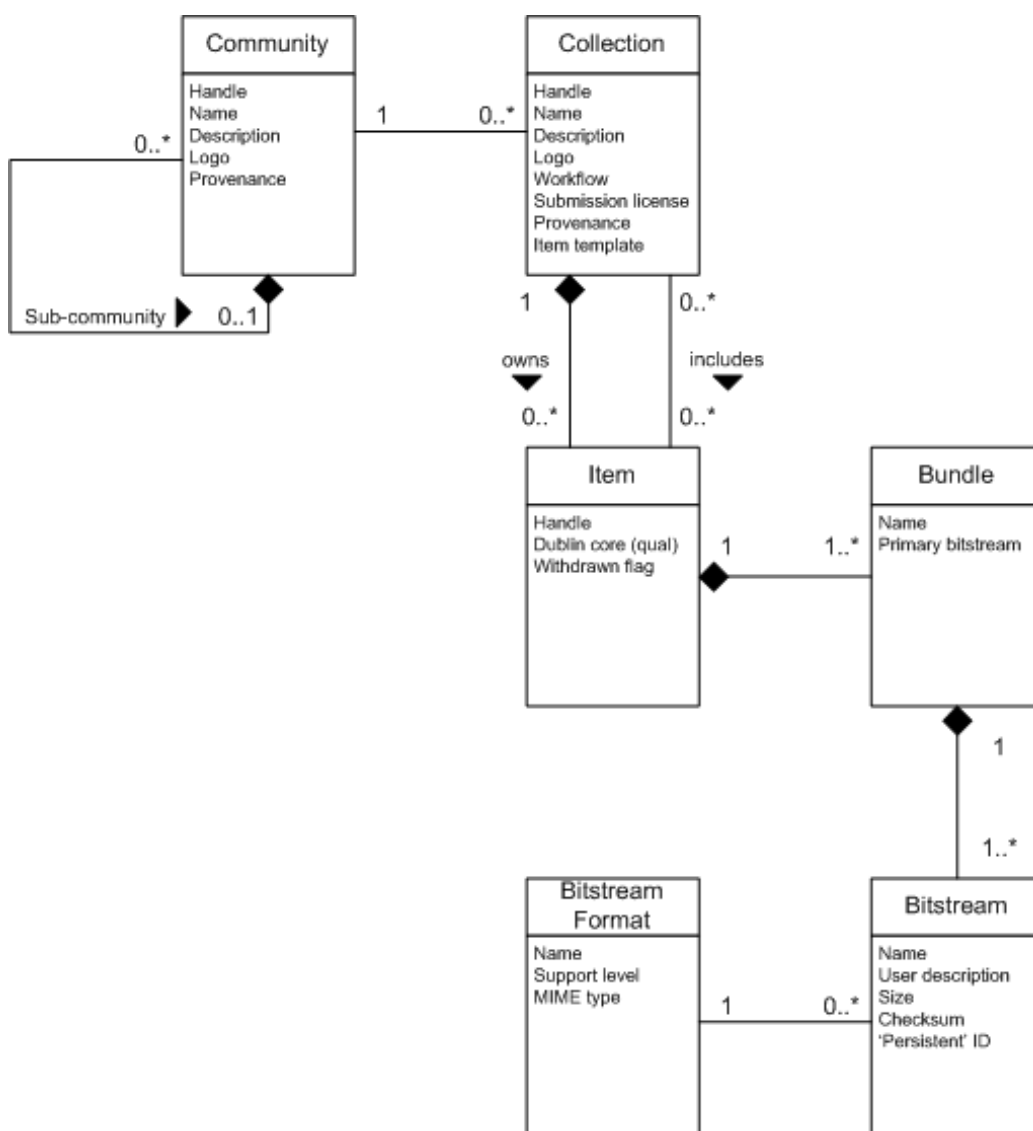


Figura 6.3: Diagrama de las entidades y sus interrelaciones.

En esta subsección se mostrarán las pruebas realizadas en la tecnología DSpace. La interface web que se utilizará es una de las que están disponibles por defecto, llamada Manakin. Un sitio DSpace está dividido en *comunidades*, las cuales pueden ser divididas en sub-comunidades reflejando una estructura universitaria de departamento, instituto, centro de investigación, o laboratorio. Las Comunidades contienen *colecciones*, los cuales son grupos de contenidos relacionados. Una colección puede aparecer en más de una comunidad. Cada colección se compone de *items*, los cuales son los elementos básicos de archivo. Cada ítem es propiedad de una colección. Adicionalmente, un ítem puede aparecer en colecciones adicionales; sin embargo cada ítem tiene una y solo una colección al cual pertenece. Los items son además subdivididos en paquetes o *bitstreams* (son flujos de bits, usualmente archivos de computadora). Los bitstreams que están relacionados fuertemente, como por ejemplo archivos HTML e imágenes que componen un único documento HTML, son organizados en paquetes. Se muestra en la figura 6.3 el diagrama de clases, en formato UML¹, para las entidades descritas. En la figura 6.4 se muestra la pantalla para el envío de un nuevo ítem en una colección seleccionada. En ella se muestran los metadatos necesarios para la creación de un nuevo ítem, de los cuales se destacan: el autor, el título, el tipo del contenido, y el lenguaje.

¹UML, (Unified Modeling Language) es el lenguaje de modelado de sistemas de software más conocido y utilizado en la actualidad; está respaldado por el OMG (Object Management Group). Es un lenguaje gráfico para visualizar, especificar, construir y documentar un sistema.

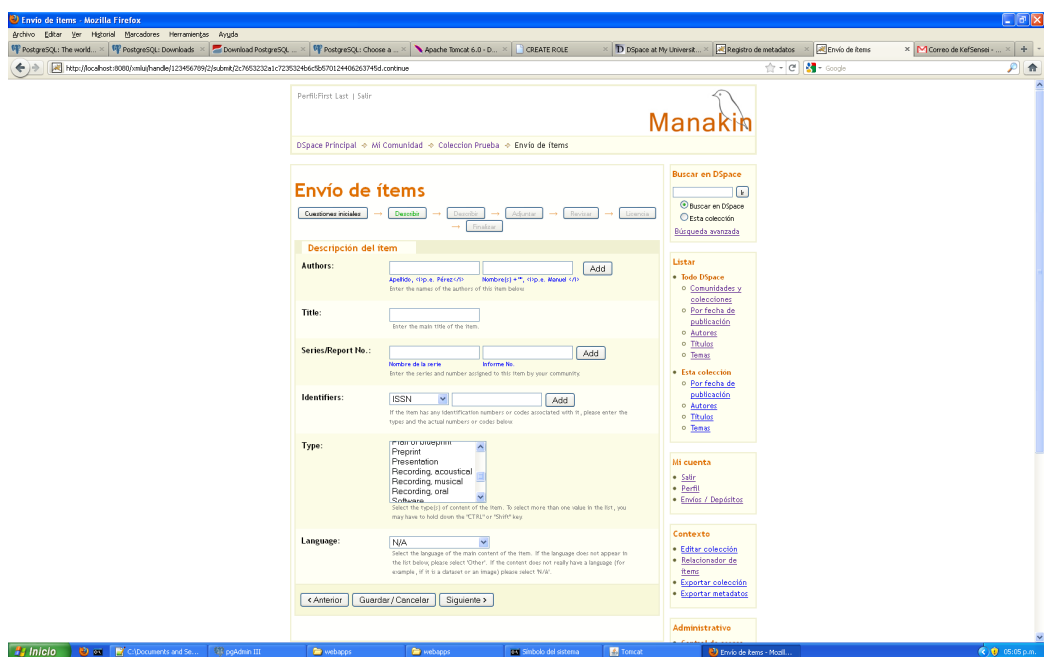


Figura 6.4: Envío de ítems

En la figura 6.4 se muestra la vista de un ítem ya creado con sus metadatos asociada. Se destaca además de los metadatos vistos para describir al ítem, una lista de ficheros que contiene, los cuales se pueden descargar para su visualización, y una lista de las colecciones a las cuales pertenece, para poder acceder a ellas de una forma simple.

DSpace Principal [Mi Comunidad](#) [Coleccion Prueba](#) [Ver ítem](#)

El Conde de Montecristo

[Mostrar el registro completo del ítem](#)

Título:	El Conde de Montecristo
Autor:	Dumas, Alejandro
Descripción:	Es un libro
URI:	http://hdl.handle.net/123456789/3
Fecha:	2011-08-20

Ficheros en el ítem

Ficheros	Tamaño	Formato	Ver	Descripción
Conde Montecristo.txt	2.840Kb	Fichero de texto	Ver/ <wbr/>Abrir	Libro de texto

Este ítem aparece en la(s) siguiente(s) colección(ones)

- [Coleccion Prueba](#)
Es una prueba

[Mostrar el registro completo del ítem](#)

Figura 6.5: Vista de un ítem dentro de una colección

En la figura 6.6 se muestra una vista del registro completo del ítem, con todos sus metadatos asociados. Una vez ingresado, los metadatos pueden ser modificados, si es que se poseen los permisos necesarios para dicho ítem. En la figura 6.7 se muestra como se puede ingresar en el campo *dc.date* (fecha) información con un formato erróneo. Para poder solucionar este problema se debería, por ejemplo, mejorar la interfaz gráfica para que no se pudiera ingresar datos con formato erróneo. No se ha intentado modificar la aplicación, pero en el sitio web de DSpace se encuentra una extensa documentación en la cual se informa sobre las distintas opciones de configuración, y en conjunto con el código fuente, se puede lograr modificar la instancia instalada, para lograr así adecuarla a las necesidades del cliente.

El Conde de Montecristo

[Mostrar el registro sencillo del ítem](#)

dc.contributor.author	Dumas, Alejandro	
dc.date.accessioned	2011-08-20T20:32:08Z	
dc.date.available	2011-08-20T20:32:08Z	
dc.date.issued	2011-08-20	
dc.identifier.uri	http://hdl.handle.net/123456789/3	
dc.description	Es un libro	en_US
dc.description.provenance	Submitted by First Last (jandres1ff@gmail.com) on 2011-08-20T20:32:07Z No. of bitstreams: 1 Conde_Montecristo.txt: 2909 bytes, checksum: f79185320580ba7a2c2859037111f4a2 (MD5)	en
dc.description.provenance	Made available in DSpace on 2011-08-20T20:32:08Z (GMT). No. of bitstreams: 1 Instalacion DSpace.txt: 2909 bytes, checksum: f79185320580ba7a2c2859037111f4a2 (MD5)	en
dc.language.iso	es	en_US
dc.subject	Conde	en_US
dc.subject	Montecristo	en_US
dc.subject	Aventuras	en_US
dc.title	El Conde de Montecristo	en_US
dc.type	Book	en_US

Ficheros en el ítem

Ficheros	Tamaño	Formato	Ver	Descripción
Conde Montecristo.txt	2.840Kb	Fichero de texto	Ver/ <wbr/>Abrir	Libro de texto

Este ítem aparece en la(s) siguiente(s) colección(ones)

- [Coleccion Prueba](#)
Es una prueba

Figura 6.6: Vista completa del registro del ítem

Aviso

New metadata was added.

Editar ítem

Estado del ítem
Archivos del ítem
Metadatos del ítem
Visualizar ítem
Curar

Añadir metadatos al ítem

Nombre:

Valor:

Idioma

ADVERTENCIA: Estos cambios no se validan de ningún modo. Es responsabilidad suya introducir los datos en el formato correcto. Si no está seguro, por favor NO haga cambios.

Metadatos

Borrar	Nombre	Valor	Idioma
<input type="checkbox"/>	dc.contributor.author	Dumas, Alejandro	<input style="width: 50px;" type="text"/>
<input type="checkbox"/>	dc.date	tyh	<input style="width: 50px;" type="text"/>

Figura 6.7: Añadir un nuevo metadato con formato erróneo

En la figura 6.8 se muestra el contenido RSS ² de una comunidad y una colección. Cualquier usuario que lo desee puede suscribirse a las actualizacio-

²RSS son las siglas de Really Simple Syndication, un formato XML para syndicar o compartir contenido en la web. Se utiliza para difundir información actualizada frecuentemente a usuarios que se han suscrito a la fuente de contenidos.[40]

nes de una colección o comunidad en particular para mantenerse informado de sus actualizaciones.

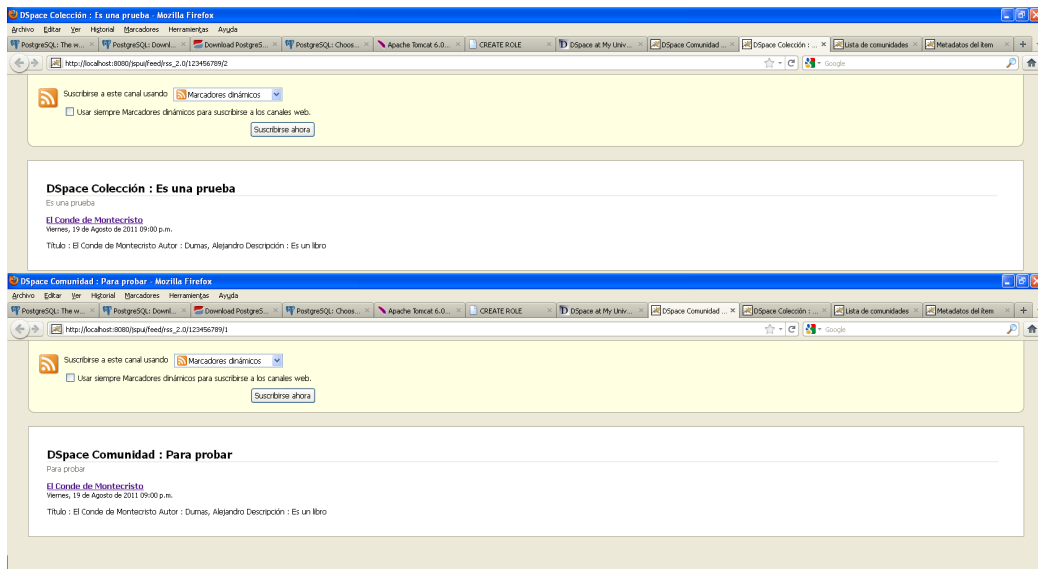


Figura 6.8: Ejemplos de RSS

En la figura 6.9 se muestra la vista de administración de privilegios avanzado. Si se es administrador de una comunidad o colección, se pueden dar o quitar privilegios a tipos de contenidos de una colección.

Administrador de privilegios avanzado

Le permite añadir y quitar privilegios a tipos de contenidos de una colección. Cuidado, es peligroso: ¡borrar el permiso LEER de algún ítem puede ocultarlos!

Para todos los grupos marcados...

Grupo:

 Administrator
 Anonymous

...concede la capacidad de llevar a cabo la siguiente acción...

Acción:

...para todos los siguientes tipos de objetos...

Tipo de contenido:

...en todas las colecciones siguientes.

Colección:

Añadir privilegios

Limpiar privilegios

Volver

Figura 6.9: Administración de privilegios avanzados para usuarios

Por último, la figura 6.10 muestra el registro de formatos de archivo, muestra una lista con información sobre los formatos de archivos conocidos y su nivel de soporte. Se puede editar o añadir nuevos formatos de archivo

con esta herramienta.

DSpace Principal [Registro de Formato](#)

Registro de formatos de archivos

Esta lista de archivos aporta información sobre formatos de archivos conocidos y su nivel de soporte. Puede editar o añadir nuevos formatos de archivo con esta herramienta. Los formatos marcados como 'internos' se ocultan al usuario, y son usados para fines administrativos.

[Añadir un formato de archivo nuevo](#)

ID	Nombre	Tipus MIME	Nivel de soporte
1	Unknown	application/octet-stream	Desconocido
<input type="checkbox"/>	License	text/plain; charset=utf-8 (<i>internal</i>)	Conocido
<input type="checkbox"/>	CC License	text/html; charset=utf-8 (<i>internal</i>)	Conocido
<input type="checkbox"/>	Adobe PDF	application/pdf	Conocido
<input type="checkbox"/>	XML	text/xml	Conocido
<input type="checkbox"/>	Text	text/plain	Conocido
<input type="checkbox"/>	HTML	text/html	Conocido
<input type="checkbox"/>	CSS	text/css	Conocido
<input type="checkbox"/>	Microsoft Word	application/msword	Conocido
<input type="checkbox"/>	Microsoft Word XML	application/vnd.openxmlformats-officedocument.wordprocessingml.document	Conocido
<input type="checkbox"/>	Microsoft Powerpoint	application/vnd.ms-powerpoint	Conocido
<input type="checkbox"/>	Microsoft Powerpoint XML	application/vnd.openxmlformats-officedocument.presentationml.presentation	Conocido
<input type="checkbox"/>	Microsoft Excel	application/vnd.ms-excel	Conocido

Figura 6.10: Registro de formatos de archivos

6.2.2. Pruebas en EPrints

En esta subsección se mostrarán algunas pruebas realizadas en la tecnología EPrints, como ingreso de un nuevo documento, modificación de los metadatos y modificaciones del template de presentación. El diseño que se muestra en estas pruebas es el que queda definido por defecto luego de la instalación.

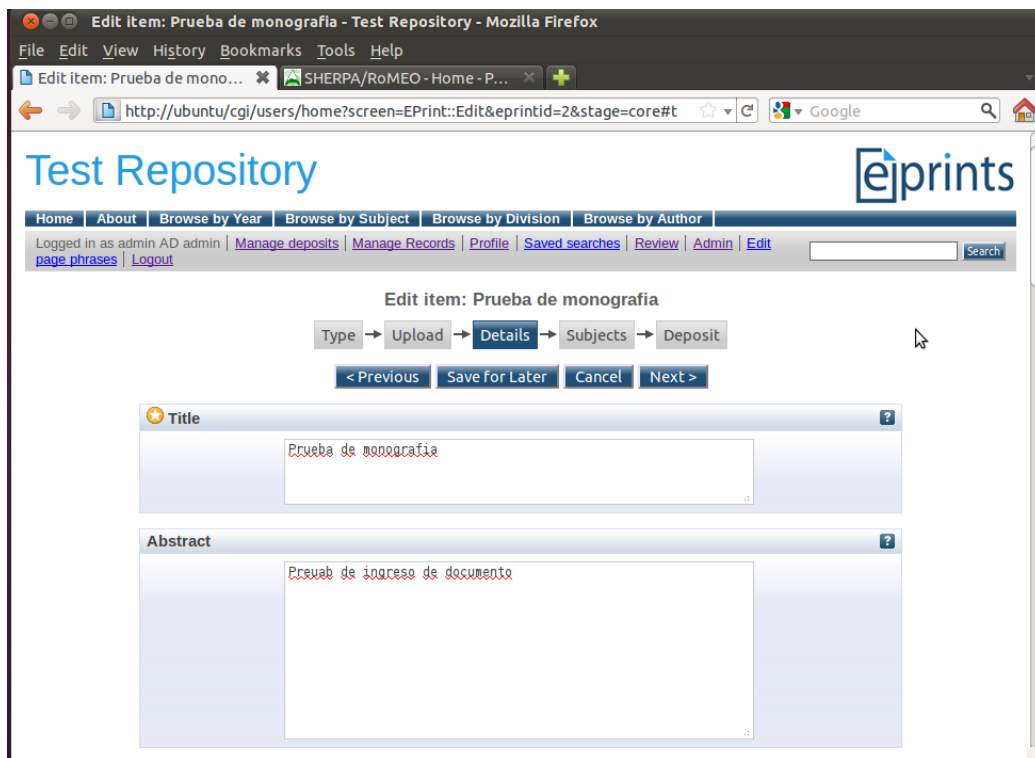


Figura 6.11: Ingreso de documento(1).

En la figura 6.11 se muestra la primera etapa de la carga de un documento donde se ingresan los metadatos de título y resumen del documento.

The screenshot shows a web browser window with the following content:

- Monograph Type:** A list of radio buttons for selecting the document type:
 - Technical Report
 - Project Report
 - Documentation
 - Manual
 - Working Paper
 - Discussion Paper
 - Other
- Creators:** A table with columns for Family Name, Given Name / Initials, and Email.

	Family Name	Given Name / Initials	Email
1.	admin	admin/ad	
2.			
3.			
4.			
- Corporate Creators:** Three empty input fields for corporate creator information.
- Divisions:** A list of pre-defined division names:
 - Faculty of Engineering, Science and Mathematics: School of Chemistry
 - Faculty of Engineering, Science and Mathematics: School of Civil Engineering and the Environment
 - Faculty of Engineering, Science and Mathematics: School of Electronics and Computer Science

Figura 6.12: Ingreso de documento(2).

En la figura 6.12 se continúa con el ingreso de metadatos de tipo, autor, división.

The screenshot shows a web browser window with the title "Edit item: Prueba de monografía - Test Repository - Mozilla Firefox". The address bar shows the URL "http://ubuntu/cgi/users/home?screen=EPrint::Edit&eprintid=2&stage=core#put". The main content area is a form titled "Publication Details" with the following fields and options:

- Status:** Radio buttons for Published (selected), In Press, Submitted, and Unpublished.
- Institution:** Text input field containing "UDELAR".
- Department:** Text input field containing "INCO".
- Place of Publication:** Empty text input field.
- Publisher:** Text input field containing "Federico".
- Identification Number:** Empty text input field.
- Number of Pages:** Empty text input field.
- Date:** Year: empty text input field; Month: dropdown menu with "Unspecified" selected; Day: dropdown menu with "?" selected.
- Date Type:** Radio buttons for UNSPECIFIED, Publication, Submission, and Completion.
- Official URL:** Empty text input field.
- Related URL:** Text input field containing "URL" and a dropdown menu for "URL Type" with "UNSPECIFIED" selected.

Figura 6.13: Ingreso de documento(3).

En la figura 6.13 se finaliza el ingreso de metadatos, agregando nombre de institución, departamento, publicador y estado. También se puede agregar numero de identificación y una fecha correspondiente al status que se eligió.

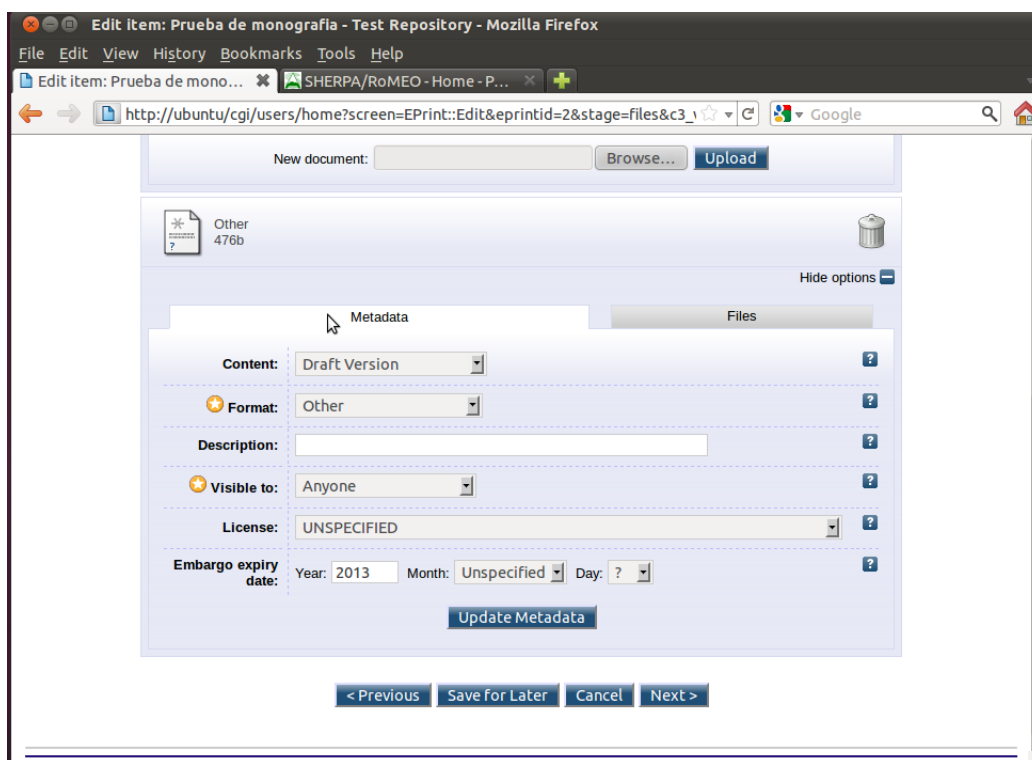


Figura 6.14: Ingreso de documento(4).

En la figura 6.14 se completa el ingreso del documento, ingresando el archivo adjunto a publicar (para esta prueba se utilizó un archivo cualquiera), se elige el formato, la visibilidad (para todos, restringida a algún grupo), licenciamiento y tiempo de validez.

SHERPA/RoMEO - Search - Publisher copyright policies & self-archiving - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Edit item: Prueba de mono... SHERPA/RoMEO - Search - P... +

http://www.sherpa.ac.uk/romeo/search.php

Google

SHERPA/RoMEO ... opening access to research

Home • Search • Journals • Publishers • FAQ • Suggest • About

English | Español | Magyar | Português

Publisher copyright policies & self-archiving

No journals found when searched for: **poblacion de uruguay**

Search again?

Journal titles or ISSNs Publisher names

Exact title starts with contains ISSN

[Advanced Search](#)

RoMEO Colour	Archiving policy
Green	Can archive pre-print and post-print or publisher's version/PDF
Blue	Can archive post-print (ie final draft post-refereeing) or publisher's version/PDF
Yellow	Can archive pre-print (ie pre-refereeing)
White	Archiving not formally supported
	More on colours and restrictions
or	View all publishers

Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

The RoMEO Journals database is supplemented with information kindly provided by:

- the British Library's [Zetoc](#) service hosted by MIMAS,
- the [Directory of Open Access Journals](#) (DOAJ) hosted by Lund University Libraries,
- the [Entrez](#) journal list hosted by the NCBI.

© 2006-2011, University of Nottingham [Contact us](#)

Figura 6.15: Revisión de derechos de autor.

En la figura 6.15 se muestra la página de validación de propiedad intelectual. Allí se puede consultar, si existieran, restricciones de publicación para el/los documentos a agregar al repositorio. Este paso no es obligatorio para el ingreso de documentos, ni siquiera aparece en la interfaz de ingreso de EPrints, sino que aparece como un link en la página de ingreso. De cualquier forma, forma parte de las buenas prácticas de ingreso de documento el chequeo de restricciones por propiedad intelectual.

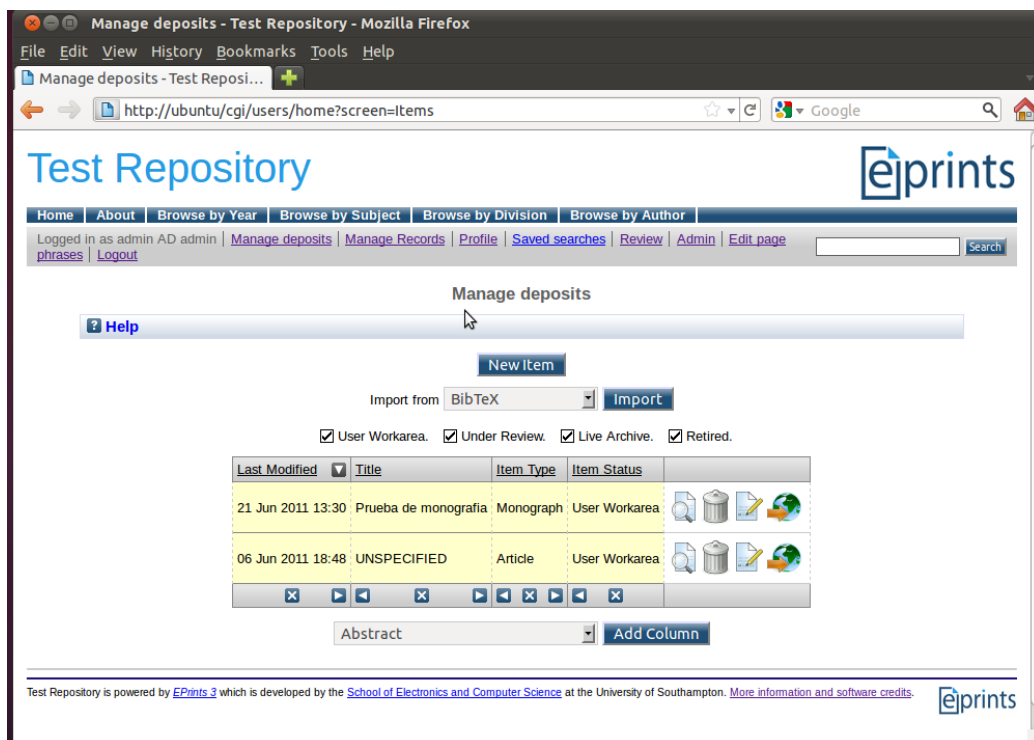


Figura 6.16: Vista principal.

En la figura 6.16 se muestra un índice general de las publicaciones del repositorio donde se incluye la recién ingresada.



Figura 6.17: Modificación de aspecto.

En la figura 6.17 se muestra la página de personalización de EPrints, donde se puede introducir cambios en las leyendas de los títulos, items, etc. En este caso modificamos el texto del nombre del repositorio “Active_name” y se agregó “Proyecto de grado”. Luego de aplicar los cambios la modificación se hace efectiva en el título del repositorio como muestra la figura.

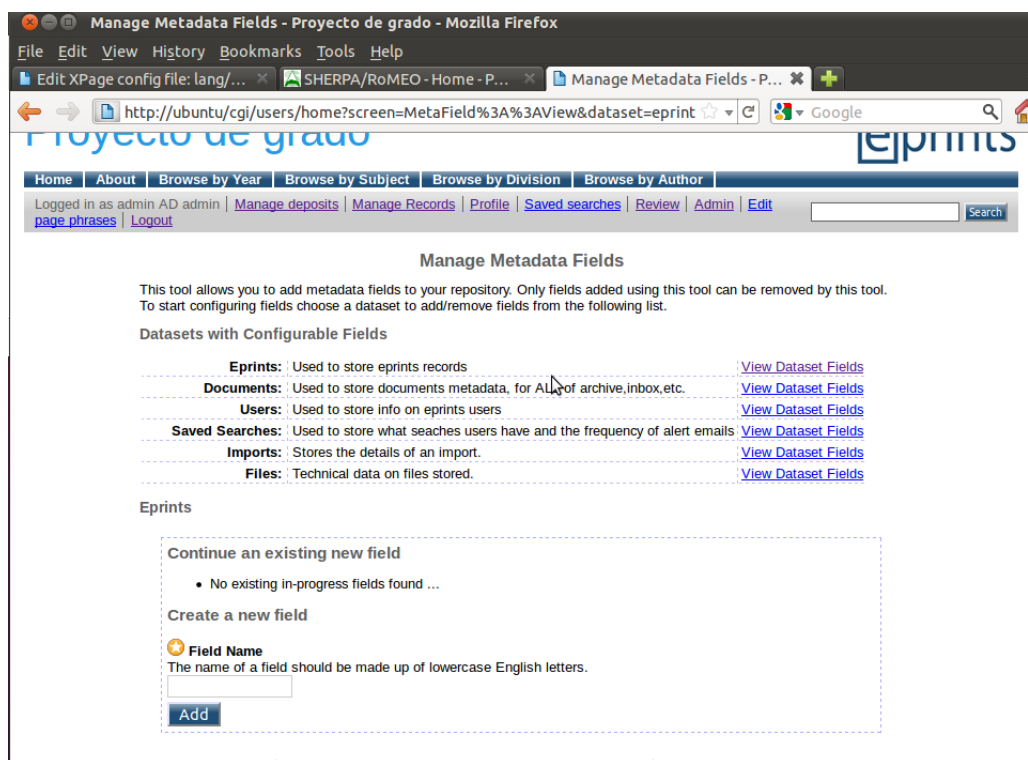


Figura 6.18: Modificación de metadatos

En la figura 6.18 se muestra la página principal de modificación de metadatos. Accediendo por ejemplo en el link correspondiente a documentos se podría agregar/modificar/quitar metadatos de ingreso como los mostrados en las figuras anteriores.

Capítulo 7

Resultados obtenidos

7.1. Evolución del proyecto

7.1.1. Descubrimiento del mundo OAI y los repositorios institucionales

Estamos viviendo en un mundo donde todo se comparte, desde fotografías y videos, hasta estados de ánimo, actividades e ideas. Un mundo donde las redes sociales están más presentes en nuestra vida diaria y donde los medios para acceder a estas redes son cada vez más accesibles en precio como teléfonos y computadoras personales y cada vez de mayor potencia.

En contraste, el acceso a publicaciones científicas, trabajos de grado y postgrado no está presente en ninguna red social (por lo menos en Uruguay!!!), y el acceso al mismo sigue restringido a las bibliotecas, o institutos de la universidad, pública o privada o en portales de acceso temático, como Timbó o SciELO restringiendo su visibilidad a un pequeño grupo de personas.

Entonces hemos descubierto a partir de este proyecto, que el problema de acceso a la información científica es resuelto en otros lugares mediante iniciativas de acceso abierto a la información (OAI) que permite conectar e intercambiar publicaciones y que éstas sean accesibles a cualquier persona que desde una computadora en cualquier lugar del mundo quiera accederla, aumentando así su visibilidad.

El acceso a esta información no sería posible si aparte de la metodología y el interés de publicar y compartir la información no hubiera software específico que permitiera el acceso. Aquí nos encontramos que existen una gran cantidad de este tipo de software, la mayoría de ellos con licencia gratuita, con una comunidad activa de usuarios muy grande y una gran cantidad de instalaciones y en algunos casos con decenas de versiones y millones de

documentos publicados y compartidos.

Concluimos que la implementación de un software de RI, unido a una metodología de publicación y utilización, sería muy útil para integrar a Uruguay a la tendencia latinoamericana y mundial de utilización de metodologías de acceso abierto, y aumentará la visibilidad de nuestra producción científica.

7.1.2. Características comunes de los RIs

Aunque existen una gran cantidad de implementaciones de RIs, durante el transcurso del proyecto fuimos descubriendo que la mayoría de las características son compartidas. Esto parece lógico si recordamos que el pilar fundamental de estos software es permitir la interoperabilidad utilizando el protocolo OAI-PMH. Para lograr interoperar se debe tener un conjunto de metadatos compatibles, por ello, cada repositorio estudiado codifica sus metadatos utilizando DUBLIN-CORE y permiten DUBLIN-CORE extendido. Además, algunos permiten la utilización de MARC21. Otra característica compartida es la autogeneración, característica que fue desarrollada anteriormente en el capítulo 4.

7.1.3. Valor agregado en la utilización de la herramienta de valoración de RI

Desde el momento que fue creada la herramienta de valoración, fue más fácil identificar cada aspecto importante a la hora de realizar una elección de software, y aportó también en la valoración de cada aspecto. De otra manera hubiera sido difícil reunir toda la información y presentarla adecuadamente. Notamos además que dándole mayor *peso* a un aspecto o grupos de aspectos podíamos incidir en los resultados logrando adaptar la herramienta a los requerimientos de usuario.

Como aspecto negativo, de tener la capacidad de incidir en los resultados, sería la pérdida de objetividad, ya que, la valoración queda muchas veces muy influenciada por la experiencia puntual obtenida en algún aspecto. Por ejemplo, experiencias de instalación o configuración que están atadas a la experiencia que tenga el técnico sobre la plataforma y el software utilizado.

Consideramos que el balance final fue positivo, y que la herramienta fue de gran utilidad en la realización del proyecto.

7.1.4. Elección de DSpace como software de repositorio

Recomendamos utilizar el software DSpace para la implementación del RI, basado en el resultado obtenido de la valoración realizada por la herramienta construida.

En general, la valoración de DSpace es superior al resto del software candidato. Este resultado se puede ver en la figura 7.1.

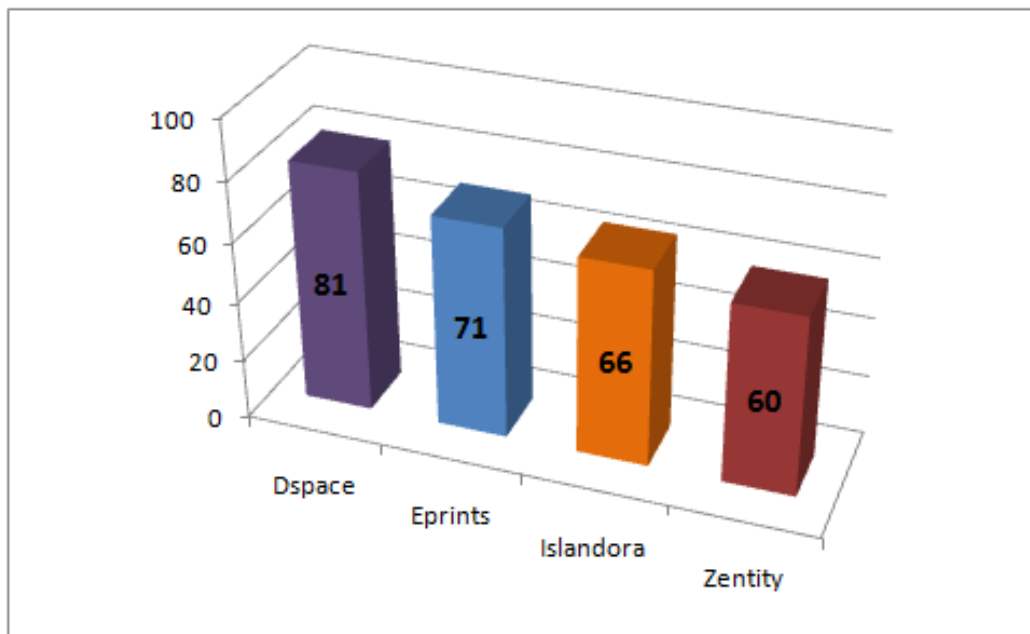


Figura 7.1: Resultado final de la valoración.

En la figura 7.2 se puede observar como se comporta cada repositorio según las categorías técnicas, funcionales y de requerimiento de usuario. Podemos observar que a nivel técnico, la valoración del software es muy similar, de esto concluimos que todos los software evaluados tienen similares características técnicas, aunque algunos se destacan en algún aspecto técnico particular y otros en otro. A nivel funcional, se puede observar un comportamiento similar al anterior. Todos el software evaluado tiene una valoración similar. Donde se aprecia una diferencia es a nivel de los requerimientos de usuario, donde DSpace tiene una valoración ampliamente superior al resto. Es por esta diferencia que recomendamos utilizar DSpace.

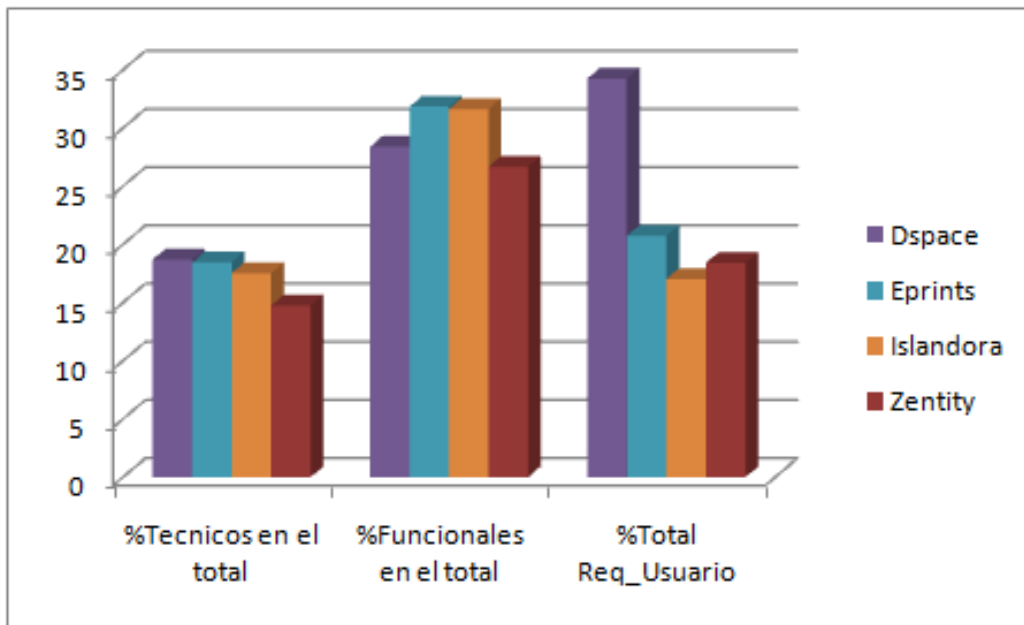


Figura 7.2: Resultados por tipo.

7.2. Cumplimiento de los requerimientos del Cliente

7.2.1. Interoperabilidad con Aleph

En el estudio realizado sobre la tecnología Aleph, se ha visto la posibilidad de interoperar entre los sistemas que utilizan Aleph y los que utilizan los repositorios estudiados, mediante el protocolo OAI-PMH. Con dicho protocolo se pueden obtener resultados de consultas hacia un sistema Aleph y mostrarlas en el sistema repositorio. De esta forma se logra tener la información de catálogo Aleph en conjunto con los datos del repositorio, centralizando las consultas de ambos sistemas, logrando así, una mayor facilidad de uso por parte de los usuarios.

Se debe tomar en cuenta que la instancia del sistema Aleph que posee la UdelaR no ha contratado el servicio para poder utilizar OAI-PMH. En consecuencia, para poder realizar la interoperación entre los sistemas con OAI-PMH, se debería contratar dicho servicio. Dado que el protocolo OAI-PMH es estándar, la comunicación entre los sistemas será más estable a cambios de versiones tanto del sistema Aleph, como del repositorio elegido. Tomando en consideración lo dicho anteriormente, se recomienda la utilización del servicio OAI-PMH para el Aleph.

7.2.2. Manejo de metadatos

Referente a las estructuras de metadatos utilizadas en los repositorios estudiados, se ha encontrado el modelo de metadatos Dublin Core. Dicho modelo es utilizado por los repositorios estudiados como su modelo de metadatos básico. Es posible extender este modelo agregando tipos de datos nuevos en cualquier momento sin la necesidad de modificar los datos ya ingresados con anterioridad. Esto lo hace muy útil al querer utilizarlo en repositorios donde se necesiten metadatos particulares, o sea necesario agregar nuevos metadatos en una forma segura respecto a los datos que ya se poseen.

7.2.3. Control de puntos de acceso

Al momento de depositar objetos y asignar sus metadatos correspondientes, surgen varias formas en las cuales poseen distintas ventajas y desventajas en su utilización. Básicamente existen dos tipos de sistemas distintos para realizar un depósito. El primero es el sistema de depósito centralizado, en el cual un grupo de personas son los encargados de realizar los depósitos e ingresar o verificar sus metadatos. Esto hace que la implementación de controles de los objetos y sus metadatos sea más sencilla, pero requiere de personal, por lo que eleva el costo de operación y manejo del repositorio. El otro sistema es el de autoarchivo, en el cual el depósito y publicación son realizados por el que desea publicarlo, lo que disminuye el trabajo de manejar el repositorio y el tiempo de publicación. Pero tiene como contraparte que la calidad del contenido es manejada por los usuarios, lo cual hace importante el crear un sistema de acreditación de usuarios para que no se pierda la calidad de los datos ingresados.

Al investigar en la utilización de repositorios institucionales para centros educativos en el mundo, se ha observado que se utiliza en su mayoría un sistema de depósito centralizado, en el cual la institución posee un mayor control sobre la calidad de los datos que son publicados en ella.

Independiente de la forma de depósito, hemos observado en diferentes implementaciones de RI que es posible definir un workflow de ingreso de datos. En la práctica se presentan como diferentes pantallas donde se muestran los metadatos a rellenar, donde algunos deberán ser ingresados con cierto formato y otros puedan ser elegidos de una lista desplegable. A modo de ejemplo, se presenta una posible secuencia de pasos de un ingreso típico.

1. Ingreso del nombre del documento, ingreso de fecha (formato específico), tipo de documento (de una lista desplegable), autor (ingreso nuevo o de una lista desplegable).

2. Ingreso del contenido del documento, mediante archivo adjunto en algún formato permitido (pdf, doc, ps, odt, etc.).
3. Universidad e instituto al que pertenece el documento (de lista desplegable).
4. Estado de documento (finalizado, borrador, etc.) (de lista desplegable).

La definición correcta del workflow aporta a mantener la calidad de los datos ingresados.

En conclusión, por los argumentos expuestos anteriormente, se recomienda la utilización de este tipo de sistemas para depositar los datos en el repositorio.

7.2.4. Experiencias en instalación y pruebas de repositorios

En las pruebas de instalación de los distintos repositorios estudiados, se ha observado que se puede realizar su instalación con relativa facilidad, aunque la dificultad varía dependiendo del sistema operativo sobre el cual se esté instalando.

En nuestras pruebas de instalación utilizamos sistemas Windows y Linux. En general, para cada repositorio a probar se debía instalar software previo, como frameworks, base de datos, servidor web independientemente del sistema operativo utilizado. Sin embargo, en el caso de Linux, algunos de los softwares son distribuidos como código fuente lo que obliga a generar los instaladores. Esto ocasionó que las instalaciones sobre Linux llevara trabajo y tiempo extra.

En este sentido, usando plataformas Windows el tiempo necesario para la puesta en producción sería menor. Como para este proyecto la plataforma debía ser software libre no se planteó la opción de utilizar un sistema operativo que necesite licenciamiento.

Con respecto a las pruebas sobre los repositorios instalados, constatamos que en todos los casos poseían un entorno gráfico (por defecto) amigable, para la administración y utilización, con herramientas de edición muy simples de utilizar y con gran cantidad de documentación y guías paso a paso, disponibles en línea.

7.2.5. Maestro de repositorios: manejo de jerarquías con el protocolo OAI-PMH

En el enfoque de RIs que cumplen con las políticas OAI, no está definida explícitamente el concepto de jerarquía. Los RIs asumen el rol de proveedores de datos (*data providers*) a los proveedores de servicios (*service providers*) que por lo general son los llamados metabuscadores. En este caso, los metabuscadores se ubicarían en el tope de la jerarquía pues serían recolectores de los metadatos, pero no recolectados por los RIs. Cabe la posibilidad que un RI sea proveedor de datos y de servicios a la vez. De esta manera podría recolectar metadatos de otros RIs y ser recolectado por otro RI o metabuscador. De esta manera se estarían agregando “niveles” que simularían una jerarquía.

Entendemos que el protocolo OAI-PMH está orientado a la inexistencia de jerarquías que restrinjan la visualización, sino que va en la línea de aumentar todo lo posible la visibilidad de los documentos.

7.3. Requerimientos no cumplidos

A continuación se detallaran los requerimientos, los cuales inicialmente figuran dentro del alcance del proyecto y no fueron cumplidos, y los motivos por los cuales no se pudieron cumplir.

7.3.1. Proponer mejoras al diseño del RI

Entendemos que uno de los motivos por el cual no se llegó a implementar una mejora puntual de alguna característica del software de RI fue que, aunque se probaron diferentes softwares, sobre diferentes plataformas, el alcance dado al proyecto no permitió que el cliente contara con una instalación de RI como para personalizar e identificar una mejora.

Aun así, teniendo en cuenta unos de los requerimientos del cliente (interoperabilidad con Aleph) una mejora que permitiría al acceso (no así la interacción) al sistema Aleph sería la implementación de una interfaz de usuario adicional que por medio de un link desde el RI fuera llamada y que esa interfaz mediante al armado de una URL específica obtuviera los datos. Ésta sería una solución rápida al acceso pero de alto costo de mantenimiento, pues cualquier cambio en el formato de la URL generaría un problema. Como se propuso anteriormente, la solución de fondo sigue siendo la compra del módulo de OAI que Aleph tiene disponible y su instalación en UdelaR.

7.3.2. Pruebas de carga

En este punto en particular no se pudieron realizar pruebas de carga sobre el producto pues durante la ejecución del proyecto y al final del mismo no se cuenta con una instalación del software de RI recomendado (DSpace), o de otro software, en un entorno de test similar en capacidad de hardware y software al entorno de producción. Además, entendemos que las pruebas de carga se deberían realizar luego de adaptar el software a las necesidades del cliente, y esta adaptación forma parte del trabajo futuro.

Preparación de pruebas de carga

Para la planificación correcta de pruebas de carga se debería tomar en cuenta los siguientes tres factores, los cuales, según la documentación y la experiencia de la comunidad de usuarios, son los que más impactan en la performance. Estos son:

1. Configuración de memoria de TOMCAT, se deberá adaptar según la memoria disponible. Se recomienda no menos de 512 MB[42].
2. Configuración de memoria de POSTGRESQL. Por defecto la instalación destina 8MB para la memoria SGA. Ésta podría aumentarse hasta 1/3 de la memoria del equipo[42].
3. Manejo de estadísticas (autocommit y warmup). Al aumentar la carga de datos también aumenta el tamaño de índices y aumenta la demora de acceso a estos datos. DSpace ofrece 2 formas de atacar este problema.

Autocommit, que implica un mejor manejo de la aplicación al impactar datos en la base de datos (no se produce espera entre la confirmación de una operación en DSpace y el comienzo de otra pues el commit es asincrónico). En las últimas versiones de DSpace, esta opción está habilitada por defecto.

Warmup en DSpace significa realizar mantenimientos periódicos de las estadísticas sobre períodos de tiempo acotados. La documentación[41] plantea modificar ciertos archivos de configuración de DSpace para mantener las estadísticas frescas de un mes de datos para mejorar la performance de acceso y creación, que según se plantea, son las operaciones más frecuentes.

Capítulo 8

Conclusiones

Durante la evolución del proyecto se logró tomar contacto con el mundo de las iniciativas de acceso abierto (OAI), los protocolos involucrados (OAI-PMH, SWORD, etc.) y el software que lo implementa. A medida que se avanzaba en el estudio del software candidato se pudo comprobar que compartían muchas de sus características técnicas y ofrecían similar funcionalidad, además de interoperar utilizando los mismos protocolos e intercambiar las mismas unidades de información, los metadatos.

En este sentido, el estudio de los metadatos involucrados, su modificación y/o adaptación a la realidad planteada por el cliente fueron unos de los puntos que a nuestro entender, se cumplieron correctamente. Otro aspecto fundamental en el proyecto, y de especial importancia para el cliente, fue el estudio de interoperabilidad de los RI, y en particular con el sistema Aleph. En este caso, se pudo realizar una recomendación, luego de haber estudiado Aleph y sus capacidades de interoperar y se plantearon alternativas en caso de que no se cuente con el modulo de Aleph necesario.

Además, se estudió la evolución cuantitativa del software de RI a nivel latinoamericano y mundial, logrando de esta manera, tener una idea clara de cuál es la tendencia de uso.

Toda la experiencia e información que se iba acumulando fue organizada en una herramienta que permitió el acceso ordenado por tipo y la valoración de cada aspecto a tener en cuenta en la elección de un software de RI. Del estudio del resultado de la herramienta se realizó la recomendación de **utilizar DSpace para la puesta en marcha del RI.**

Se valoró como muy positiva su utilización, pues permitió compactar en pocas páginas una gran cantidad de información que de otra manera hubiera sido difícil de manejar, y permitió la correcta ponderados de los aspectos

a evaluar. La ponderación de aspectos se realizó en cooperación con los técnicos de EUCBA, que aportaron su conocimiento.

Las pruebas de instalación del software de RI candidato sobre diferentes sistemas operativos nos brindaron la posibilidad de enfrentar los posibles problemas que se pueden dar en un ambiente de producción y nos permitió interactuar con la herramienta, experiencia muy útil a la hora de valorar aspectos técnicos y dificultad de instalación y configuración.

8.1. Trabajos a futuro

Consideramos que la puesta en marcha del RI utilizando el software seleccionado es uno de los puntos fuertes del trabajo futuro, pues aunque el software provee gran parte de la funcionalidad necesaria, es necesario trabajar, y mucho, sobre la personalización de la interfaz gráfica y la creación y/o modificación de estructuras y metadatos para adecuar el software a las necesidades del cliente.

También entendemos que es necesario generar un cambio cultural en los actores que generan el conocimiento científico y de las autoridades educativas en general, el cual incentive a la publicación de sus trabajos en un RI destinado a ese fin. Creemos que el proyecto de EUCBA contribuye a lograrlo con la creación de políticas claras a la hora de realizar publicaciones de material científico.

Otro aspecto que aumentaría la visibilidad de la producción científica sería la integración de las redes sociales actuales al RI. Las funcionalidades que poseen las redes sociales de, por ejemplo, comentar lo publicado en ellas, lleva a que se pueda conocer la opinión de las personas, conocidas o no, acerca de las publicaciones. Esto además puede ser útil para preguntar sobre dudas de la publicación comentada, o sobre otras publicaciones que puedan ser del mismo tipo en el repositorio de la publicación u otro que alguno de los usuarios conozca. Las posibilidades que aportan las redes sociales para el intercambio de información, son en consecuencia, un buen motivo como para tener en cuenta en la puesta en marcha de un RI.

Glosario

DC Dublin Core.

DCMI Dublin Core Metadata Initiative.

EUBCA Escuela Universitaria de Bibliotecología y Ciencias Afines.

framework En el desarrollo de software, un framework o infraestructura digital, es una estructura conceptual y tecnológica de soporte definido, normalmente con artefactos o módulos de software concretos, con base a la cual otro proyecto de software puede ser más fácilmente organizado y desarrollado. [45].

interoperabilidad Según NISO (National Information Standards Organization): habilidad de múltiples sistemas, con diferentes plataformas de hardware, software, estructuras de datos e interfaces, de intercambiar y compartir datos.

MARC MACHine-Readable Cataloging.

metadatos Pueden ser definidos como “datos de los datos” que describen el contenido, calidad, condición y otras características de los datos. Los metadatos son vitales para ayudar a los usuarios potenciales a encontrar los datos buscados y para determinar si este conjunto de datos satisfacen sus necesidades antes de procesarlos (con el gasto asociado en tiempo y dinero).

OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting.

Open Archives Initiative Es uno de los movimientos internacionales que fomenta el desarrollo y promueve la el uso de un protocolo estándar (Open Archives Metadata Harvesting Protocol (OAMHP)) diseñado para mejorar la búsqueda y recuperación de “e-prints” que residen en repositorios distribuidos. Además promueve la creación de estándares de interoperabilidad que facilite la disseminación de contenidos.

recolectar En el contexto de OAI (“Open archive initiative”, “Iniciativa de acceso abierto”), “harvesting” refiere específicamente a la reunión de metadatos desde repositorios distribuidos en un repositorio centralizado.

RI repositorio institucional.

RIs repositorios institucionales.

script En informática, un guión, archivo de órdenes o archivo de procesamiento por lotes. Es un programa usualmente simple, que por lo regular se almacena en un archivo de texto plano. [46].

SeCIU Servicio Central de Informática.

SHERPA/RoMEO RoMEO es una base de datos de búsqueda de políticas de publicación que es utilizada en operaciones de autoarchivo en la web, y por repositorios OAI compliant. [47].

SWORD Simple Web-service Offering Repository Deposit.

UdelaR Universidad de la República.

Referencias

- [1] *XLII Jornadas Mexicanas de Biblioteconomía*,
<http://www.ambac.org.mx/jornadas/XLII/data/uploads/15.pdf>,
última visita realizada el 17/AGO/2011.
- [2] *Datos sobre el acceso abierto en España*,
<http://www.biblogtecarios.es/davidgomez/datos-sobre-el-acceso-abierto-en-espana-el-caso-de-los-repositorios-institucionales>, última visita realizada el 3/FEB/2012.
- [3] *Wiki de EPrints*, <http://wiki.eprints.org>, última visita realizada el 6/SEP/2011.
- [4] , <http://eprints.ucm.es>, última visita realizada el 6/SEP/2011.
- [5] *Página de DSpace*, <http://www.dspace.org>, última visita realizada el 03/SEP/2011.
- [6] *Detalles de instalación DSpace*: en el archivo comprimido descargado para realizar la instalación del producto, en la carpeta: “dspace-1.7.2-release\dspace\docs\pdf”, o en la página de dspace <https://wiki.duraspace.org/display/DSDOC/DSpace+Documentation>, última visita realizada el 07/JUN/2011.
- [7] *Ejemplos de casos de uso de DSpace*,
<http://www.dspace.org/use-cases/international-survey-of-institutional-digital-repositories-published.html>, última visita realizada el 03/SEP/2011.
- [8] *Página principal de Zentity*,
<http://research.microsoft.com/en-us/projects/zentity/>,
última visita realizada el 17/AGO/2011.

- [9] *Guía de instalación Zentity*, archivo “Zentity Version 2 - Installation Instructions.docx” incluido en el descargable en la página, última visita realizada el 17/AGO/2011.
- [10] *Windows SDK for Windows 7*,
<http://www.microsoft.com/downloads/details.aspx?FamilyID=6b6c21d2-2006-4afa-9702-529fa782d63b&displaylang=en>,
17/AGO/2011.
- [11] *.Net Framework 4*,
<http://www.microsoft.com/downloads/details.aspx?FamilyID=9cfb2d51-5ff4-4491-b0e5-b386f32c0992&displaylang=en>, última visita realizada el 17/AGO/2011.
- [12] *Silverlight 4*, <http://www.microsoft.com/getsilverlight/Get-Started/Install/Default.aspx>, última visita realizada el 17/AGO/2011.
- [13] *Windows PowerShell 2.0*, <http://www.microsoft.com/downloads/en/details.aspx?FamilyID=50633A1F-A665-425D-923C-1A269F8AC084&displaylang=en>, última visita realizada el 17/AGO/2011.
- [14] *Budapest Open Access Initiative*,
<http://www.soros.org/openaccess/read>, última visita realizada el 10/AGO/2011.
- [15] *La urgente necesidad de construir Repositorios Institucionales*, presentación realizada por Laureano Gómez Dueñas,
<http://www.slideshare.net/bibliounad/la-urgente-necesidad-de-construir-repositorios-institucionales>, última visita realizada el 16/NOV/2011.
- [16] *OAIster*, <http://oaister.worldcat.org/>, última visita realizada el 16/NOV/2011.
- [17] *Acceso a base de datos OAIster*,
<http://www.oclc.org/oaister/access/default.htm>, última visita realizada el 16/NOV/2011.
- [18] *The Directory of Open Access Repositories (Open DOAR)*,
<http://www.openoar.org>, última visita realizada el 20/SEP/2011.

- [19] *Registry of Open Access Repositories*, <http://roar.eprints.org/>, última visita realizada el 20/SEP/2011.
- [20] *Aleph Integrated Library System*, <http://www.exlibrisgroup.com/category/Aleph>, última visita realizada el 16/NOV/2011.
- [21] *BiUR: Bibliotecas de la Universidad de la República*, <http://www.biur.edu.uy/F>, última visita realizada el 02/NOV/2011.
- [22] *Biblioteca Nacional Uruguay*, <http://www.bibna.gub.uy/>, última visita realizada el 02/NOV/2011.
- [23] *Open Archives Metadata Harvesting: An Overview*, <http://shodhganga.inflibnet.ac.in/dxml/bitstream/handle/1944/1130/15.pdf?sequence=1>, última visita realizada el 02/NOV/2011.
- [24] *SECIU*, <http://www.seciu.edu.uy/>, última visita realizada el 02/NOV/2011.
- [25] *Dublin Core*, <http://dublincore.org/>, última visita realizada el 08/DIC/2011.
- [26] *Publishing-Based OAI Data Provider*, *Publishng_Based_OAI_Data_Provider.pdf*, última visita realizada el 08/DIC/2011.
- [27] Lynch, C., *Intitutional Repositories: Essential Infraestructure for Scholarship in the Digital Age*. ABL Bimonthy Report 226, 2003.
- [28] Crow, R., *The Case for Institutional Repositories: A SPARC Position Paper*. Washington DC, Scholary Publishing and Academic Resources Coalition: 37 2002.
- [29] Heery, R. and S. Anderson, *Digital Repositories Review, UKOLN and AHDS: 33*. 2005.
- [30] *Web 2.0*, http://es.wikipedia.org/wiki/Web_2.0, última visita realizada el 25/FEB/2012.
- [31] *Streaming*, <http://es.wikipedia.org/wiki/Streaming>, última visita realizada el 25/FEB/2012.

- [32] *Definición de SWORD*, <http://www.arvo.es/dspace/sword/>, última visita realizada el 25/FEB/2012.
- [33] *Avro Consultores*, <http://www.arvo.es/>, última visita realizada el 25/FEB/2012.
- [34] *Estudio DSpace por Avro*, <http://www.arvo.es/dspace/>, última visita realizada el 04/MAR/2012.
- [35] *Estructura de los archivos de importación de items*, <http://www.arvo.es/dspace/estructura-de-los-archivos-de-importacion-de-items/>, última visita realizada el 04/MAR/2012.
- [36] *XMLUI o JSPUI??*, <http://www.arvo.es/dspace/xmlui-o-jspui/>, última visita realizada el 04/MAR/2012.
- [37] *SWORD Word Add-in*, <http://research.microsoft.com/en-us/projects/authoring/>, última visita realizada el 25/FEB/2012.
- [38] *DRIVER Guide lines Spanish*, http://www.driver-support.eu/documents/DRIVER_2_0_Guidelines_Spanish.pdf, última visita realizada el 25/FEB/2012.
- [39] *OpenAIRE*, <http://www.openaire.eu/es/open-access/open-access-overview>, última visita realizada el 25/FEB/2012.
- [40] *RSS Wikipedia*, <http://es.wikipedia.org/wiki/RSS>, última visita realizada el 25/FEB/2012.
- [41] *Increasing DSpace performance*, <https://atmire.com/website/?q=content/increasing-dspace-performance>, última visita realizada el 28/MAR/2012.
- [42] *Performance Tuning DSpace*, <https://wiki.duraspace.org/display/DSDOC18/Performance+Tuning+DSpace>, última visita realizada el 28/MAR/2012.
- [43] *DSpace AIHT*, http://www.digitalpreservation.gov/partners/documents/DSpace_AIHT.pdf, última visita realizada el 28/MAR/2012.

- [44] *Scalability Issues*, <https://wiki.duraspace.org/display/DSPACE/ScalabilityIssues>, última visita realizada el 29/MAR/2012.
- [45] *Framework Wikipedia*, <http://es.wikipedia.org/wiki/Framework>, última visita realizada el 30/MAR/2012.
- [46] *Script Wikipedia*, <http://es.wikipedia.org/wiki/Script>, última visita realizada el 30/MAR/2012.
- [47] *SHERPA/RoMEO*, <http://www.sherpa.ac.uk/romeo/faq.php?fIDnum=1&mode=simple&la=en#what-is>, última visita realizada el 01/ABR/2012.