



Macroproyecto
Tecnologías para la Universidad
de la Información y la Computación
U N A M



Red de Repositorios Universitarios
de Recursos Digitales

Etapa 1: Investigación

Primer informe técnico

Responsables:

Clara López Guzmán – DGSCA (Coordinadora)

Alberto Arriaga Arredondo - DGSCA

Alberto Castro Thompson – DGB

Isabel Galina Russell – DGSCA

Fernando Gamboa Rodríguez – CCADET

Joaquín Giménez Heau - Instituto de Biología

Pablo Miranda Quevedo – DGSCA

Agosto, 2006

CONTENIDO

1.	INTRODUCCIÓN.....	5
2.	LOS REPOSITORIOS.....	7
2.1	Tipos de repositorios.....	8
2.1.1	Los repositorios de eprints y temáticos	8
2.1.2	Los repositorios de materiales académicos	9
2.1.3	Los repositorios institucionales	10
2.2	La visibilidad	11
2.3	La interoperabilidad	11
2.4	Reforma en la comunicación y publicación académica	12
3.	INICIATIVAS	13
3.1	Open Access	13
3.2	Iniciativas de proyectos conjuntos	16
3.2.1	Dawning of the Dutch Network of DARE (Digital Academic REpositories).....	16
3.2.2	Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe CLACSO 17	
3.2.3	Harvesting Institutional Resources in Scotland Testbed (HaIRST).....	17
3.2.4	Proyecto Sherpa.....	18
3.2.5	MIT OCW (OpenCourseWare).....	19
3.2.6	eScholarship Repository	20
3.2.7	SciDev.Net	20
3.2.8	BioMed Central.....	21
3.3	Índices y cosechadores de repositorios	21
4.	TECNOLOGÍAS PARA LA CREACIÓN Y OPERACIÓN DE RI.....	23
4.1	Esquema de Metadatos Dublin Core.....	24
4.2	El protocolo OAI-PMH.....	29
4.3	Herramientas de software	32
4.3.1	CDSware.....	32
4.3.2	DSpace	33
4.3.3	Fedora (Flexible Extensible Digital Object and Repository Architecture)	33
4.3.4	Eprints	34
5.	INTERFACES DE USUARIO EN SISTEMAS DE BÚSQUEDA.....	35
5.1	El papel de la interfaz	35
5.2	El diseño de interfaces para la búsqueda de información.....	36
5.3	El humano como procesador de información.....	38
5.4	El modelo cognitivo distribuido.....	39
5.5	El problema.....	39

6.	ANÁLISIS DE CASOS	41
6.1	Metodología	41
6.2	Resultados	42
6.2.1	Software	43
6.2.2	Tipo de repositorio.....	46
6.2.3	Políticas de ingreso.....	48
7.	LOS REPOSITORIOS EN LA UNAM	49
7.1	Los repositorios universitarios.....	49
7.2	Selección de repositorios candidatos	50
7.3	Instituto de Biología- UNIBIO	52
7.4	BIDI-MADI.....	52
7.5	BIDI-UNAM	53
7.6	Instituto de Investigaciones Sociales (IIS).....	54
7.7	Colecciones mexicanas – DGSCA.....	55
7.8	Centro de Investigaciones en Energía (CIE).....	55
8.	REFLEXIONES FINALES Y TRABAJO FUTURO	57
9.	REFERENCIAS.....	59
10.	BIBLIOGRAFÍA RECOMENDADA.....	63
11.	ANEXOS	69
	ANEXO A. Requerimientos técnicos de Software.....	69
	ANEXO B. Dublin Core Qualified.....	72

1. Introducción

El proyecto 3R-Red de Repositorios Universitarios de Recursos Digitales¹, forma parte del Macroproyecto de Tecnologías para la Universidad de la Información y la Computación. Es un proyecto que consta de 4 etapas: Investigación, Modelo Conceptual, Desarrollo e Implementación.

El objetivo del proyecto está encaminado a la creación del prototipo de una red de repositorios de la UNAM, que permita mayor uso y visibilidad de la producción intelectual de los miembros de la comunidad.

Este informe técnico presenta los resultados de la primera etapa del proyecto, en la que se realizó una investigación exploratoria durante 4 meses, para conocer el estado del arte de los repositorios a nivel mundial y poder ubicar la situación de la UNAM en este ámbito.

La primera parte de la investigación consistió en conocer a detalle qué es un repositorio y los distintos tipos que se han dado origen. Posteriormente, se realizó un análisis de diversas iniciativas que están dando soporte al desarrollo de los repositorios; estas iniciativas tienen que ver tanto con tendencias en el desarrollo de contenidos como proyectos líderes a nivel mundial. También se hace una revisión a las principales tecnologías que se están utilizando para crear repositorios estándares e interoperables, así como las herramientas de software que son ampliamente utilizadas para la creación de repositorios. Se plantea, por otra parte, una reflexión sobre la problemática de las interfaces de usuario en sistemas de búsqueda, en la que se sitúa la relevancia del diseño de interfaces y la problemática de transformar al ser humano a un procesador de información.

Con todo el antecedente teórico sobre el tema de repositorios, se procedió a hacer un análisis de casos, en el que se estudió a un conjunto de repositorios para tener dimensiones sobre el software que utilizan, los tipos de repositorios más comunes y algunos comportamientos en sus políticas de ingresos de

¹ <http://www.3r.unam.mx/>

contenidos. Una vez realizado este estudio de repositorios internacionales, se procedió a hacer un diagnóstico de la situación de los repositorios dentro de la UNAM, para lo cual, con base en todo lo anteriormente investigado, primeramente se tipificó un repositorio universitario y considerando estos criterios, aún generales, se comentan algunos proyectos de la UNAM que podrían ser potenciales repositorios y candidatos para el prototipo de 3R.

Finalmente, se presentan algunas reflexiones sobre el tema de repositorios y de cómo se incorporan estos modelos al contexto del manejo de contenidos, acorde a los requisitos y operación de las dependencias Universitarias.

2. Los repositorios

Actualmente los repositorios son una herramienta para administrar y difundir los recursos electrónicos producidos por miembros de diversas comunidades, incrementando y fortaleciendo el acceso a los recursos académicos a nivel institucional y mundial. Sin embargo, no existe un consenso general sobre las características exactas que deben tener estos depósitos de contenidos para ser considerados como un repositorio, ello debido en gran parte a que existen divergencias en torno a los objetivos que se quieren lograr con la creación y mantenimiento de un repositorio en una institución académica. Esto conlleva a discusiones en torno a quiénes son responsables de construir y mantener los repositorios, el tipo de material que se puede depositar, la asignación de derechos patrimoniales, la cobertura del costo de instalación y mantenimiento, así como los mecanismos para asegurar la calidad, integridad y preservación del material (Chan, 2004; Ware, 2004).

Un repositorio no es sólo una colección de objetos digitales y es común confundirlos con bibliotecas digitales², la línea es muy fina entre ambos, por ello es importante dejar claro qué es un repositorio.

De acuerdo a Heery et. al (2005), existen cuatro principales características de un repositorio: la primera es que debe contar con mecanismos que permitan el depósito de material por parte del creador, el dueño u otra persona (por ejemplo, un bibliotecario); la arquitectura del repositorio debe manejar tanto el contenido como sus metadatos; deben existir servicios básicos como búsqueda y recuperación, administración, controles de acceso y permisos, entre otros. Por último, el repositorio debe de ser sustentable a largo plazo, administrado y

² Por biblioteca digital se entienden colecciones de recursos digitales que tienen un objetivo claro, formadas por una selección de contenidos organizados con un sistema descriptivo a través de metadatos (catalogación), y que además tienen asociadas algunas facilidades para la búsqueda y uso de la información (servicios), usualmente los recursos son adquiridos a algún tercero.

apoyado de forma seria por parte de una organización confiable (Heery et. al 2005).

Por lo tanto, un repositorio no se refiere únicamente a los contenidos ya que la forma de administración, sus funciones y servicios forman parte de sus características. Para Lynch (2003) un repositorio es 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.

Idealmente, un repositorio es una iniciativa institucional y está apoyado por las estructuras administrativas correspondientes, tanto a nivel general como particular³.

2.1 Tipos de repositorios

En la práctica, se identifican diferentes tipos de repositorios, algunos dependen del tipo de validación de sus contenidos, algunos otros por la forma en que se administran. A continuación se describen algunos de éstos.

2.1.1 Los repositorios de eprints y temáticos

Los primeros repositorios se crearon para que los investigadores pudieran colocar sus artículos electrónicos en línea y para que otros investigadores del mismo campo pudieran acceder a ellos a través de las interfaces de búsqueda y recuperación. El motivo principal era poder compartir sus resultados de investigación de una forma más rápida, mejorando así la comunicación entre los investigadores. Adicionalmente, de esta forma se eliminaban las barreras económicas de acceso que existen para consultar los artículos en revistas

³ Cabe aclarar que el actual consenso en las evaluaciones de proyectos de repositorios es que los principales retos que enfrentan los repositorios son de tipo cultural, organizacional y administrativo, más que de índole tecnológico (Ware 2004, Awre et.al 2005).

académicas comerciales. (Harnard, 2001).

Estos repositorios generalmente se conocen como temáticos ya que son creados y mantenidos por una institución, pero cualquier persona puede depositar material de un tema en particular. Generalmente, son repositorios que albergan artículos científicos (*eprints*), ya sea la versión previa a la publicación (*preprint*), la versión publicada o una versión posterior (*post-print*). Un ejemplo es el repositorio de arXiv, mantenido por Los Alamos, el cual contiene *eprints* de las áreas de física, matemáticas, ciencias de la computación y biología cuantitativa.

El éxito de estos repositorios se debe a que fue uno de los principales motores para el movimiento de acceso abierto u *Open Access* (OA). El movimiento propone que los resultados de la investigación científica, en forma de artículos arbitrados, estén disponibles digitalmente en línea y de forma gratuita, así como libre de prácticamente toda restricción de copiado (*copyright*) (Suber 2004). Los repositorios de este tipo tienden a solamente aceptar artículos arbitrados y de ciertos temas.

OA intenta eliminar barreras económicas como pagos de suscripción a revistas, así como barreras de permiso, principalmente de reproducción y distribución, a la literatura científica mundial. Este movimiento se describe con mayor detalle en la sección 3.1.

2.1.2 Los repositorios de materiales académicos

Existe un modelo de repositorios que se considera como una herramienta útil para ofrecer una amplia gama de materiales académicos y no sólo de artículos arbitrados. Se parte del hecho de que muchos repositorios, pero no necesariamente todos, apoyan el movimiento de *Open Access*, ofreciendo artículos arbitrados en su colección (Heery et. Al, 2005) pero también contenidos de valor académico que no pasan por este proceso.

Este modelo aumenta las posibilidades de acceso a los recursos electrónicos, permite generar nuevas formas de publicación y arbitraje, facilita compartir y reutilizar datos crudos de investigación u objetos de aprendizaje, entre otras

funciones (Heery et. al, 2005). Uno de los principales objetivos es apoyar y mejorar la enseñanza, aprendizaje e investigación en la institución. (Chan, 2004).

Este tipo de repositorios ofrecen materiales diversos y no sólo de artículos científicos en forma de *eprints*. Se podría crear, por ejemplo, un repositorio de objetos de aprendizaje ⁴ o de especímenes biológicas.

2.1.3 Los repositorios institucionales

Este tipo de repositorios, que incluyen material académico diverso, tienden a ser organizados por una institución más que por áreas temáticas. El objetivo principal es que el repositorio funciona como un tipo de vitrina para mostrar la producción académica de la institución que lo maneja (Andrew, 2003). Se conocen como repositorios institucionales (RI) (*institutional repositories*).

Los RI están basados y apoyados en una institución académica y ofrecen material digital producido por sus miembros. Tanto el contenido del repositorio como las políticas de selección y almacenamiento de los materiales está definido institucionalmente (Johhson, 2002). Aunque las características adicionales pueden variar, Ware (2004) observa que todos los repositorios institucionales tienen por lo menos las siguientes características en común:

- son accesibles en línea y contienen material académico que está definido y producido por una institución;
- el repositorio es interoperable y abierto al utilizar un software para el intercambio de metadatos con otros repositorios, que colecciona, almacena y difunde el material.

En este sentido los repositorios institucionales forman parte del proceso de comunicación académica y tienen el compromiso de ser acumulativos y

⁴ Una unidad con un objetivo didáctico, caracterizada por ser digital, independiente, con una o pocas ideas relacionadas, accesible a través de metadatos con la finalidad de ser reutilizadas en diferentes contextos y plataformas de aprendizaje en línea.

perpetuos.

Actualmente, hay una tendencia mundial en incorporar repositorios a las instituciones académicas, algunas ya cuentan con él y otras están en vías de hacerlo.

2.2 La visibilidad

Un común denominador para los repositorios de todo tipo es el fin último de incrementar la visibilidad⁵ y uso de los documentos electrónicos ahí depositados. Ware (2004) revisó el contenido de cuarenta y cinco RI y descubrió que el 58% del material disponible eran documentos varios, mientras que el 22% eran *eprints* y el 20% restante consistía de tesis.

Se han hecho estudios en torno a los repositorios de *eprints* y parece existir una tendencia a que se incremente el número de citas de un *eprint* en comparación con un impreso (Lawrence, 2001). Se considera por lo tanto, que los repositorios pueden ayudar a resolver el problema de "invisibilidad" que padecen muchos de los recursos electrónicos que están en línea. Los repositorios pueden contribuir a que la producción académica de una institución tenga mayor visibilidad, mejorando la comunicación académica así como el impacto del trabajo realizado.

2.3 La interoperabilidad

Independientemente del tipo de repositorio, la interoperabilidad⁶ entre los distintos repositorios que se están creando, es una característica fundamental para que se tenga el máximo impacto deseado.

La iniciativa de archivos abiertos *Open Archives Initiative (OAI)* junto con su protocolo para cosechar metadatos (*OAI-Protocol Metadata Harvesting*, ver apartado 4.1), han sido tecnologías claves para el desarrollo de los repositorios

⁵ Entendiendo por visibilidad que un recurso electrónico obtenga lectores e impacto deseados y medibles.

⁶ Condición mediante la cual sistemas heterogéneos o de naturaleza distinta pueden intercambiar procesos o datos.

a nivel mundial, ya que permite de forma sencilla que los diversos repositorios pueden ser buscados y la información recuperada.

2.4 Reforma en la comunicación y publicación académica

Hoy día, existen debates importantes en torno a la relación entre los repositorios institucionales y temáticos. Al margen del las mismas, ambos tipos de repositorios generalmente son vistos como importantes estrategias por parte de las universidades para acelerar cambios en la comunicación y publicación académica (Ware, 2004).

Se considera que los repositorios forman parte esencial de la infraestructura para promover cambios y mejoras en las formas de comunicación y publicación académica, que ahora son posibles a partir de las nuevas tecnologías de información. Así mismo, forman parte importante de las estrategias consideradas para el futuro en términos de ciberinfraestructura, tanto para las Ciencias como para las Humanidades (Hey 2005; Unsworth 2005; Atkins 2003); aunque como se ha comentado, sus funciones específicas continúan debatiéndose. Independientemente de esto, es claro que los repositorios en sus diversas formas anteriormente descritas continuarán evolucionando y jugando un rol cada vez más importante en la comunicación y publicación académica.

3. Iniciativas

El desarrollo del campo de los repositorios ha puesto en marcha diversas iniciativas que buscan su estandarización tecnológica y operativa. También hay iniciativas o movimientos que están revolucionando los esquemas tradicionales de la publicación y edición de publicaciones científicas. En este apartado se hace una revisión general a éstas.

3.1 *Open Access*

El fruto del trabajo intelectual de investigadores en todo el mundo ha estado "monopolizado" por las casas editoras de revistas de divulgación científica debido a que son la herramienta para poder dar a conocer los resultados de su trabajo. Como consecuencia la difusión del conocimiento se ha visto limitado tanto por los altos costos que involucra la publicación tradicional, como por la monopolización antes mencionada. Aquí se presenta una gran paradoja: los científicos e investigadores que pertenecen a organizaciones financiadas por fondos públicos, y que no pueden pagar los altos costos de suscripción a revistas de literatura científica, no pueden acceder a sus propios trabajos, siendo ellos los que producen y ofrecen gratuitamente la materia prima de la que se nutren las revistas. Por lo que ofrecen gratuitamente su trabajo, porque no buscan una recompensa económica inmediata sino el prestigio académico en base al reconocimiento de sus publicaciones.

OA surge como un movimiento que cuestiona el monopolio que las grandes editoriales ejercen sobre la distribución de la información científica y propone dos soluciones para que la literatura erudita sea más accesible:

1. Que los autores publiquen en revistas OA ⁷ó
2. Que los autores además de publicar en revistas científicas, depositen una copia de su artículo en un repositorio temático o institucional para que otras personas puedan consultarlo.

⁷ En el sitio de DOAJ, <http://www.doaj.org/>, se encuentra un listado revistas bajo OA.

El movimiento OA consiste, de acuerdo con la Declaración de Budapest⁸ (*Budapest Open Access Initiative*), en que cualquier usuario pueda leer, descargar, copiar, distribuir, imprimir, con la posibilidad de buscar o enlazar todos los textos de artículos de literatura erudita, recorrerlos para indexación exhaustiva, usarlos como datos para software, o utilizarlos para cualquiera otro propósito legal, sin barreras financieras, legales o técnicas, distintas de la fundamental de tener acceso a la propia Internet. La única limitante a la reproducción y distribución de los artículos publicados, y la única función del copyright en este dominio, no puede ser otra que dar a los autores control sobre la integridad de su trabajo y el derecho a ser apropiadamente acreditados y citados.

La Declaración de Berlín⁹ (Acceso Abierto al Conocimiento en Ciencias y Humanidades) nació para promover Internet como el instrumento funcional que sirva de base global del conocimiento científico y la reflexión humana, todo ello bajo OA. Adicionalmente, asienta las pautas a seguir para una mejor difusión y distribución del conocimiento científico en base a la satisfacción de dos condiciones:

1. El (los) autor(es) y depositario(s) de la propiedad intelectual de tales contribuciones deben garantizar a todos los usuarios por igual, el derecho gratuito, irrevocable y mundial de acceder a un trabajo erudito, lo mismo que licencia para copiarlo, usarlo, distribuirlo, transmitirlo y exhibirlo públicamente, y para hacer y distribuir trabajos derivados, en cualquier medio digital para cualquier propósito responsable, todo sujeto al reconocimiento apropiado de autoría (los estándares de la comunidad continuarán proveyendo los mecanismos para hacer cumplir el reconocimiento apropiado y uso responsable de las obras publicadas, como ahora se hace), lo mismo que el derecho de efectuar copias impresas en pequeño número para su uso personal.

⁸ <http://www.soros.org/openaccess/read.shtml/>

⁹ http://www.mpg.de/pdf/openaccess/BerlinDeclaration_en.pdf/

2. Una versión completa del trabajo y todos sus materiales complementarios, que incluya una copia del permiso del que se habla arriba, en un conveniente formato electrónico estándar, se deposita (y así es publicado) en por lo menos un repositorio online, que utilice estándares técnicos aceptables, que sea apoyado y mantenido por una institución académica, sociedad erudita, agencia gubernamental, o una bien establecida organización que busque implementar el acceso abierto, distribución irrestricta, interoperabilidad y capacidad archivística a largo plazo.

Siguiendo los pasos de las iniciativas de Budapest y Berlín, está la carta de ECHO¹⁰ que tiene como objetivo la definición de criterios para la explotación adecuada de las potencialidades de los nuevos medios para la preservación archivística, la exploración académica y la distribución pública del patrimonio cultural de la humanidad.

Un cambio tan drástico como el propuesto por el movimiento OA genera oposición en muchos sectores pero tal y como dice Antonio Lafuente (2004): "el conocimiento debería ser patrimonio de todos, un bien común que legaremos a las generaciones futuras de la misma forma en la que nosotros hemos recibido las leyes de la gravitación de Newton, las de circulación de la sangre de Harvey, la Teoría de la Relatividad de Einstein, el principio de Arquímedes o el mapa del genoma humano. Lo cierto es que con la llegada de las nuevas tecnologías de la información los problemas han adquirido una magnitud portentosa y las alarmas han saltado por todos los rincones. Negar el acceso a la información científica a los ciudadanos, ya sea por su condición de supuestos iletrados, ya sea porque su familia, institución o país no dispone de recursos suficientes, es una opción por completo inaceptable que retrasa el avance de la ciencia y cuestiona los fundamentos mismos de la democracia."

OA está abriendo un camino para aquellos científicos que desean difundir sus conocimientos e investigaciones, sin ánimo de lucro y en algunos casos de

¹⁰ *European Cultural Heritage Online*, <http://echo2.mpiwg-berlin.mpg.de/home/project/pilotphase>

forma directa, sin la intervención de editoriales. Las instituciones tienen en OA una opción para concentrar y poseer la producción de sus investigaciones, para difundirse y aprovecharse por más comunidades.

3.2 Iniciativas de proyectos conjuntos

Además de las iniciativas internacionales en las que se firma un tratado o acuerdo para aumentar la visibilidad y el impacto de la investigación a través de Internet, numerosas organizaciones a nivel mundial se han dado a la tarea de impulsar los repositorios con la producción de la investigación que generan sus universidades e investigadores, sumándose a diferentes iniciativas de trabajo colaborativo que han surgido alrededor del mundo y que plantean soluciones aplicables a situaciones específicas.

Entre los proyectos más difundidos están:

3.2.1 Dawning of the Dutch Network of DARE (Digital Academic REpositories)

<http://www.ariadne.ac.uk/issue41/vanderkuil/>

DARE es una iniciativa común de universidades holandesas cuyo objetivo es hacer que su producción académica sea digitalmente accesible. También colaboran la Biblioteca Nacional de los Países Bajos, la KNAW (Academia de Artes y Ciencias de los Países Bajos) y la NWO (Organización de los Países Bajos para la Investigación Científica). Este proyecto comenzó en Enero de 2003 y su culminación esta planeada para diciembre de 2006.

DARE tiene varias metas, entre las que destacan:

- La implementación de la infraestructura básica para el funcionamiento e interacción de repositorios.
- Estimular el desarrollo de servicios basados en la información que se obtiene a partir de las investigaciones que están disponibles a través de la infraestructura mencionada en el punto anterior.
- Iniciar y promover el envío de contenido científico hacia los repositorios y el uso del mismo.

El programa DARE esta financiado con fondos públicos, a través del Plan de

Acción Nacional para la autopista electrónica (NAP, National Action Plan electronic highway). Con esta concesión el gobierno holandés está dando un impulso ejemplar a la innovación en la disposición de la información académica en los Países Bajos.

3.2.2 Red de Bibliotecas Virtuales de Ciencias Sociales de América Latina y el Caribe CLACSO

<http://www.clacso.org.ar//biblioteca>

La red CLACSO, a la que se puede tener acceso por Internet y sin ningún costo, tiene como objetivo principal el promover y facilitar el acceso a los resultados de las investigaciones de los centros que son miembros.

Entre los servicios que proporciona esta Red de Bibliotecas sobresalen la Sala de Lectura con textos completos de libros, artículos, ponencias y documentos de trabajo publicados por la red CLACSO y otras instituciones; Bases de datos sobre la producción académica de los centros miembros: registros bibliográficos de las publicaciones, investigaciones con descripción de cada investigación, investigadores con dirección de correo electrónico de contacto; enlaces a bibliotecas y bases de datos de ciencias sociales.

3.2.3 Harvesting Institutional Resources in Scotland Testbed (HaIRST)

<http://hairst.cdldr.strath.ac.uk>

Es uno de 14 proyectos iniciados por JISC (*Joint Information System Committee*) dentro del programa *Focus on Access to Institutional Resources* (FAIR).

Este proyecto comenzó en agosto de 2002 por un consorcio de tres universidades escocesas: Strathclyde vía el Centro para la Investigación de la Biblioteca Digital (CDLR), Napier y St.Andrews. HaIRST enfoca su investigación en el diseño, la puesta en práctica y el despliegue de un servicio experimental para un acceso autónomo, a nivel todo Reino Unido, de recursos institucionales creados en Escocia. El principal objetivo del proyecto es investigar y aconsejar

sobre algunos de los requisitos o requerimientos técnicos, culturales, y de organización asociados al depósito, al acceso, y al descubrimiento de recursos institucionales en el ambiente de la información.

Uno de los miedos a la total implementación del modelo OA es que se produzca un descenso en la calidad de las publicaciones dadas las libertades que plantea el modelo, a saber: la ausencia de un filtro editorial que evalúe y de visto bueno a las publicaciones, mecanismo con el que sí cuentan las revistas de literatura científica de prestigio. Debido a lo anterior HaIRST entiende que a nivel cultural y de organización, se necesita asegurar la existencia de los ambientes institucionales que estimulan y sostienen la creación y el depósito de recursos de calidad, y de las políticas de desarrollo de colecciones colaborativas que apoyan la actividad cooperativa en el área.

3.2.4 Proyecto Sherpa

<http://www.sherpa.ac.uk/>

Al igual que HaIRST, Sherpa formaba parte del programa JISC FAIR y se dedicaba a investigar cuestiones a futuro en cuanto a comunicación académica y a desarrollar Repositorios Institucionales Open Access en varias universidades para facilitar la difusión mundial rápida y eficiente de la investigación.

Entre las metas de este proyecto resaltan:

- Poner en marcha RI Open Access de *eprints* que utilicen el protocolo OAI-PMH y que utilicen el software Eprints (ver apartado 4.3.4).
- Investigar cuestiones claves en la creación, población y mantenimiento de colecciones de *eprints*, incluyendo: propiedad intelectual, control de calidad, políticas de desarrollo de colecciones, modelos de negocios, cultura de comunicación académica y estrategias institucionales.
- Trabajar con los proveedores de servicios OAI para alcanzar estándares aceptables (técnicos, en metadatos y de manejo de colecciones) para la difusión eficaz del contenido.
- Investigar la preservación digital de *eprints* usando el modelo de

referencia *Open Archival Information System (OAIS)*

- Divulgar sus experiencias y proporcionar consejos a otros que desean instalar servicios similares.

El proyecto finalizó en enero de 2006, pero mucho de su trabajo en defensa y ayuda en el establecimiento de RI está continuando con el proyecto SHERPA Plus.

3.2.5 MIT OCW (OpenCourseWare)

<http://ocw.mit.edu/index.html>

Este proyecto entró en funcionamiento, en una fase piloto en septiembre de 2002. Es una fuente libre y abierta de recursos docentes (*open educational resource* -- OER) y de estudios puestos a libre disposición por el MIT (*Massachusetts Institute of Technology*) en Internet dirigidos a educadores, estudiantes y autodidactas alrededor del mundo. MIT OpenCourseWare (MIT OCW) apoya la misión del MIT: avanzar en el conocimiento y la educación y servir al mundo en el siglo XXI.

El MIT cuenta con organizaciones internacionales que traducen y proveen la infraestructura necesaria para que los materiales educativos que proporcionan se encuentren disponibles en múltiples idiomas logrando con esto un impacto mucho mayor.

Entre los colaboradores del MIT OCW se pueden mencionar:

- Universia.- Consorcio de más de 800 colegios de España, Portugal, y América Latina que han traducido cerca de 100 cursos al español y al portugués.
- *China Open Resources for Education (CORE)*.- Consorcio de universidades líder en China que han traducido 115 cursos al chino simplificado.
- *OpenSource OpenCourseWare Prototype System (OOPS)*.- Organización de voluntarios que ha traducido 27 cursos al chino tradicional.

Adicionalmente existen cursos traducidos al tailandés, francés, alemán, vietnamita, y ucraniano.

3.2.6 eScholarship Repository

<http://www.cdlib.org/programs/escholarship.html>

Algunas de las principales actividades de la Universidad de California es asegurar la creación, diseminación y preservación de los productos de la investigación y la enseñanza; en base a su experiencia ha llegado a la conclusión de que la comunicación es la base para todas las actividades académicas al interior de una organización educativa y se ha dado cuenta, al igual que otras organizaciones que se han sumado a la iniciativa OA, que la publicación de materiales académicos se ha vuelto costosa y de acceso restringido. Por ello, por medio de la biblioteca Digital de California patrocina la Iniciativa eScholarship. Los servicios que ofrece esta iniciativa son en respuesta a la necesidad de encontrar mecanismos alternativos de publicación proporcionando acceso persistente y que el contenido pueda encontrarse con facilidad.

eScholarship Repository es una infraestructura libre y de acceso abierto que ofrece a los departamentos, centros y unidades de investigación de la Universidad de California control directo sobre la creación y diseminación de su producción intelectual tanto de investigación como de enseñanza, haciéndolos disponibles a todo público de manera online.

3.2.7 SciDev.Net

<http://www.scidev.net/>

Originalmente la Red de Ciencia y Desarrollo (*Science and Development Network* -- SciDev.Net) fue puesta en funcionamiento por personal de la *Journal Nature*, con ayuda financiera de la *Wellcome Trust* del Reino Unido, para difundir lo que acontecía en la Conferencia Mundial sobre Ciencia que se llevo a cabo en Budapest en 1999. La idea tuvo una excelente aceptación al grado de proponer, y después de un tiempo llevar a la realidad, la creación de un sitio web permanente de acceso libre en donde convergieran las visiones de redes regionales de trabajo de individuos e instituciones que dieran pie a la publicación de acontecimientos importantes, información confiable, opiniones y

análisis sobre ciencia y tecnología.

SciDev.Net apunta a proporcionar información confiable y de buenas fuentes acerca de la ciencia y la tecnología en un mundo en desarrollo. La meta de esta red es ayudar tanto a organizaciones como a individuos en países en vías de desarrollo para tomar decisiones informadas sobre cómo la ciencia y la tecnología pueden mejorar el desarrollo económico y social.

3.2.8 BioMed Central

<http://www.biomedcentral.com/>

Es una casa editorial independiente comprometida a proporcionar acceso abierto inmediato a investigación biomédica. En esta iniciativa, el temor de que la calidad de los artículos publicados mediante el modelo OA desaparece ya que BioMed Central esta comprometida a mantener altos estándares por medio de una completa y rigurosa revisión de los documentos que publica, adicionalmente ofrece una amplia variedad de publicaciones y otros servicios siempre conservando la mirada en una misma dirección: que todos los artículos publicados por esta casa editorial sean de libre acceso y puedan ser re-utilizados y re-distribuidos. Principios básicos que rigen el movimiento OA.

3.3 Índices y cosechadores de repositorios

La proliferación mundial de Repositorios ha dado pie a la creación de índices que son servicios que facilitan la detección e interoperabilidad de los repositorios. Los dos principales índices de RI son ROAR y OpenDOAR. ROAR (Registry of Open Access Repositories), de la Universidad de Southampton, cuenta con más de 730 repositorios registrados y permite encontrar repositorios por país, tipo de material y software utilizado. Este servicio será complementado por el nuevo servicio de OpenDOAR, de la Universidad de Nottingham, el cual reúne los repositorios OA del mundo. El sistema de búsqueda permite localizar repositorios por país, tipo de material, tema, así como realizar búsquedas libres.

Por otro lado existen los cosechadores de repositorios que permiten a los

usuarios buscar en un conjunto de repositorios con una sola petición de búsqueda y utilizando un punto de acceso común. Un ejemplo de cosechador de repositorios es OAIster, un proyecto de la Biblioteca Digital de la Universidad de Michigan en conjunto con la Universidad de Illinois, que hace la cosecha sobre los metadatos de recursos electrónicos de más de 650 RI. Se pueden buscar recursos electrónicos académicos por palabra clave, autor, título, tema y lenguaje, así como ver la lista de repositorios cosechados. Con esto, busca crear una colección de recursos digitales académicos y de difícil acceso haciéndolos de fácil consulta para cualquiera.

4. Tecnologías para la creación y operación de RI

Para su implementación y desarrollo, los RI requieren de la interrelación de varias tecnologías que cubren distintos ámbitos de funcionamiento; esta sección se centra en tres de los componentes principales relacionados: el primero aborda la utilización de los metadatos de acuerdo a su estructuración e intercambio, el segundo se refiere al protocolo OAI-PMH utilizado para su transmisión e interoperabilidad, y el tercero a la revisión del software comúnmente utilizado en los sistemas generales de repositorios.

De cualquier manera, las tecnologías específicas en un RI pueden ser varias e implementarse en distinta forma y para ello se necesita un marco tecnológico que resuelva y articule los requerimientos particulares de la institución que lo establece, así como las adaptaciones necesarias y específicas de cada comunidad académica. Previamente y antes de puntualizar en los tres aspectos mencionados anteriormente, nos referiremos de manera general y básica a las características que deben cubrir en un repositorio estas tecnologías (Barton, Waters 2004-2005; Jones et.al, 2006):

- Que tengan soporte comunitario
- Fácilmente integrables
- Que proporcionen mecanismos de autenticación y autorización
- Que proporcionen mecanismos de verificación y seguridad de los contenidos
- Que tengan un sistema de administración central
- Que establezcan claramente las licencias de contenido y sus restricciones
- Servicios web habilitados
- Que utilice uno o varios esquemas flexibles de metadatos (descriptivos, técnicos, de preservación, derechos) para la captura y encapsulado de la información de los archivos
- Que soporten la federación y escalabilidad de los repositorios
- Mecanismos de captura y egreso

- Mecanismos y políticas de preservación digital, respaldo y recuperación de datos
- Internacionalización y soporte en varios idiomas
- Que maneje distintos formatos de archivos (texto, imágenes, conjuntos de datos, video, audio, simulaciones, etc.)
- Que sea interoperable: compatibilidad con OAI, Z39.50, SRW, etc.

Asimismo se debe pensar en una solución tecnológica con posibilidades de integración a los servicios de información locales preexistentes en la institución, es necesario que estos sistemas se ajusten dentro de las infraestructuras previas en su marco de comunicación y funcionamiento. La arquitectura general del software adoptado debe ser modular para permitir adaptaciones departamentales sin menoscabo del sistema general. Con un sistema adecuado a distintos niveles es posible encapsular componentes en otros sistemas y proveer interfaces de servicios web que hagan que las tecnologías aplicadas al repositorio puedan participar en las redes distribuidas del funcionamiento general.

De forma amplia, la plataforma tecnológica de repositorios institucionales consiste de los siguientes elementos:

- Servidores Windows o Unix/Linux
- Un servidor web, como Apache y sus herramientas de aplicación web
- Un manejador de bases de datos como MySQL, DB2, Oracle, Post gres, SQL Server
- Dirección URL permanente y persistente

4.1 Esquema de Metadatos Dublin Core

Los metadatos se refieren a la información acerca de un recurso digital y facilitan la descripción y recuperación del mismo en los sistemas de información, su manejo en los repositorios es la clave para abrir el contenido digital de las instituciones y hacerlos accesibles, para estos la descripción de un objeto -quién lo creó, cómo se llama, de qué se trata, cuándo fue publicado, etcétera- es una de las dimensiones más importantes para que los contenidos puedan ser

registrados y expuestos. De acuerdo con su función, existen distintos tipos de metadatos:

- Metadatos descriptivos: tienen propósito de descubrimiento (cómo se encuentra un recurso), identificación (cómo un recurso puede distinguirse de otro), y selección (cómo determinar que un recurso cubre una necesidad particular). Los metadatos descriptivos sirven también para formar colecciones de recursos similares. Otras funciones de los metadatos descriptivos son la evaluación, relación (con otros recursos) y usabilidad
- Metadatos administrativos: es información que facilita la administración de los recursos. Incluyen información sobre cuándo y cómo fue creado el recurso, quién es el responsable del acceso o de la actualización del contenido y también se incluye información técnica, como la versión de software o el hardware necesario para ejecutar dicho recurso.
- Metadatos estructurales: sirven para identificar cada una de las partes que componen al recurso, definen la estructura que le da forma. Por ejemplo, un libro, que contiene capítulos y páginas, se puede etiquetar con metadatos que identifican cada parte y la relación que guardan entre ellas. Se usan especialmente para el procesamiento de la máquina y por software de presentación o estilos.

En estas categorías de metadatos y para diferentes tipos de contenidos existe esquemas de metadatos específicos a ciertas áreas de conocimiento, o mas bien, a cierto tipo de recursos con particularidades propias de su campo de uso. La elección de un esquema¹¹ particular de almacenamiento de metadatos depende del tipo de contenidos almacenados, formatos y prácticas de la institución que provee el almacenaje, por esta razón es indispensable que los repositorios utilicen los estándares más comunes y flexibles para producir un

¹¹ Los esquemas de metadatos consisten en un conjunto de reglas semánticas, sintácticas y de contenido que deben seguirse para conformar el conjunto de metadatos de un recurso.

conjunto de metadatos que pueda ser colectado apropiadamente.

Los metadatos varían notablemente en su complejidad, alcance y contenido. La selección correcta de dicho esquema, la asignación adecuada de sus valores y la tecnología compatible entre los sistemas dan a los contenidos las propiedades necesarias para potenciarlos como recursos reutilizables, asequibles y durables (López, 2005).

Los RI utilizan el esquema de metadatos llamado Dublin Core para describir el contenido de sus objetos, su uso se generalizó porque fue especificado por el protocolo OAI-PMH como el "requisito mínimo" para soportar cosecha de datos e interoperabilidad¹² en un repositorio estructurado.

La iniciativa del conjunto de metadatos Dublin Core fue creada con el objetivo de facilitar la detección de recursos electrónicos en línea. La Iniciativa de Metadatos Dublin Core (DCMI)¹³ es una organización dedicada a la promoción y difusión de normas interoperables sobre metadatos y el desarrollo de vocabularios especializados en metadatos para la descripción de recursos que permitan sistemas de recuperación más inteligentes. El resultado de la primera reunión en 1995 fue un conjunto de elementos de metadatos que facilitan esta labor. Actualmente es la iniciativa más extendida en Internet y es considerada un estándar (ISO-15836-2003). El Dublin Core se destaca por su deliberada sencillez. Pretende ofrecer a los creadores de materiales digitales en general una lista de descriptores sencillos, fáciles de entender y que no requieran de amplios conocimientos en catalogación para utilizarlas.

El conjunto del Dublin Core consta de 15 elementos, aunque ninguno es

¹² La interoperabilidad se define por la capacidad de un sistema de repositorios institucionales para proveer acceso a través de múltiples motores de búsqueda y otras herramientas, permitiendo mantener y exponer su metadatos a otros servicios de cosecha y búsqueda de contenido. Para muchas instituciones esta simplicidad facilita las barreras en la operación del repositorio y sólo requiere un sistema de archivos para agrupar el contenido y la habilidad para crear y compartir su metadatos con sistemas externos (Crow, 2002)

¹³ Dublin Core Metadata Initiative, <http://es.dublincore.org/about/>

obligatorio y todos son repetibles. Se pueden clasificar en tres tipos: los que tienen que ver con el contenido del recurso, los referentes a la propiedad intelectual y los relacionados con la creación e identidad del recurso (Ver Tabla 1).

Contenido	Propiedad intelectual	Creación e identidad
Título	Creador	Fecha
Tema	Editor	Tipo
Descripción	Colaborador	Formato
Fuente	Derechos	Identificador
Lengua		
Relación		
Cobertura		

Tabla 1. Elementos de Dublin Core.

Para una descripción detallada del funcionamiento y los elementos que componen el esquema Dublin Core, consultar el Anexo B.

La simplicidad del Dublin Core permite mayor participación en la creación de metadatos y promueve la interoperabilidad. Sin embargo, es limitada en su manejo de complejidad semántica y no tiene los alcances que otros estándares, tales como el MARC¹⁴, pueden manejar.

Se considera que el Dublin Core privilegia visibilidad en detrimento de la complejidad. El Dublin Core puede servir como punto de partida para la

¹⁴ El formato MARC (MACHINE Readable Cataloguing) usado masivamente para catalogar libros y revistas de acuerdo a un esquema muy detallado se convirtió en un estándar internacional, si bien MARC puede ser usado para la descripción de objetos distintos a los libros y revistas es poco apropiado para describir los elementos que integran el registro de objetos digitales, ha sido paulatinamente sustituido en los Repositorios Institucionales por otros más apropiados y flexibles.(Jones et. al. 2006)

creación de descriptores más complejos. Está diseñado específicamente para interactuar con otros esquemas. De esta forma los elementos básicos del Dublin Core pueden ser reutilizados para esquemas más complejos o algunos elementos básicos de esquemas complejos pueden ser trasladados sin dificultades al Dublin Core. Adicionalmente existe un Dublin Core Calificado que utiliza refinadores para que el significado de un elemento sea más estrecho o más específico.

El añadir nuevos esquemas no es un asunto trivial, ciertas decisiones oportunas de diseño permiten que estos se incorporen con un alto grado de modularidad y un sistema de mapeo de elementos internos muy robusto, permitiendo al conjunto central de metadatos ser expuesto de la misma forma que otros esquemas, esta conjunción puede ser ejemplificada por la manera en la cual los repositorios establecen sus estándares internos de metadatos y los transforman en contenedores que cumplen con el formato de Dublín Core soportado por OAI-PMH.

En este sentido y gradualmente la multidimensionalidad de los objetos digitales ha generado nuevos estándares de metadatos que proveen una variedad de esquemas más adecuados a las diferentes dimensiones de los objetos (descriptiva, administrativa y estructural junto con el contenido primario), comúnmente referidos como, *complex objects formats* (formatos de objetos complejos), estas iniciativas utilizan una estructura de datos basada en XML, las más avanzadas e importantes son: METS (*Metadata Encoding Transmisión Standard*); MPEG-21 DIDL (*Digital Item Declaration Language*); IEEE LOM (*Learning Object Metadata*)¹⁵.

¹⁵ METS (Metadata Encoding Transmisión Standard), <http://www.loc.gov/standards/mets/>, Desarrollado por OCLC (Online Computer Library Center) y la Biblioteca del Congreso de Estados Unidos; MPEG-21 DIDL (Digital Item Declaration Language), Las especificaciones de este estándar no están tan fácilmente disponibles en línea, pero un panorama general de éste se puede consultar en <http://www.dlib.org/dlib/november03/bekaert/11bekaert.html>,

En conclusión, los Repositorios Institucionales pueden utilizar y soportar distintos esquemas de metadatos, interpretarlos y aplicarlos en una manera estandarizada a un rango heterogéneo de formatos de objetos digitales y adecuarse a las necesidades descriptivas que generen sus comunidades académicas. Sin embargo, para la interoperabilidad será requisito que dichos esquemas sean compatibles o transformables al esquema Dublin

4.2 El protocolo OAI-PMH

El protocolo para la transmisión de contenidos en Internet denominado OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*) [<http://www.openarchives.org>] tuvo sus primeros trabajos de desarrollo desde 1999. Originalmente pensado para mejorar el acceso a los repositorios de *eprints*, depósitos de documentos de investigación científica a texto completo, fue paulatinamente mejorado y adaptado hasta convertirse en el estándar utilizado para facilitar la disponibilidad de cualquier tipo de documento digital procedente de distintos repositorios, por medio de la transmisión de metadatos a través de la World Wide Web, utilizando estándares abiertos HTTP (*Hypertext Transport Protocol*) y XML (*eXtensible Markup Language*).

La premisa básica de funcionamiento de OAI-PMH es que ofrece un marco para la interoperabilidad de repositorios independiente de la plataforma o aplicación que utilicen, mediante la codificación de sus metadatos en el esquema Dublin Core¹⁶. La simplicidad en el uso e implementación de este esquema soluciona los problemas derivados de las conversiones entre múltiples formatos al convertir sus datos a un formato común. Dublin Core utiliza 15 elementos básicos y se establece como el requisito mínimo que permite la interoperabilidad y recuperación en los RI que utilizan metadatos simples y multidisciplinares.

Desarrollada por Los Alamos National Laboratory Digital Library; Learning Object Metadata (LOM), <http://ltsc.ieee.org/wg12/>, Desarrollado por IEEE Learning Technology Standards Committee.

¹⁶ <http://es.dublincore.org/>

A diferencia de otros protocolos con una funcionalidad más compleja como el Z39.50¹⁷, OAI-PMH es intencionalmente más simple y fácil de implementar y permite niveles de interoperabilidad que posibilitan hacer peticiones de información de una manera estructurada y manejable. Junto con los protocolos gemelos SRU (*Search/Retrieve via URL*) y SRW (*Search Retrieve Web Service*), OAI-PMH es de los más usados. Es útil para el repositorio planteado habilitar su soporte para uno o más de estos de protocolos

La arquitectura de OAI-PMH se basa en clientes y servidores. Los primeros son los proveedores que proporcionan la información (metadatos) para ser cosechados por los segundos, proveedores de servicios (búsqueda y recuperación) o recolectores; crean y ofrecen herramientas de acceso a los metadatos cosechados, tales como interfases de búsqueda, estadísticas de tipos, accesos, cantidad, etc. Los repositorios institucionales generalmente tienen ambos papeles ofreciendo sus metadatos para cosechar así como servicios de consulta para el usuario final.

OAI-PMH utiliza transacciones HTTP para emitir preguntas y obtener respuestas entre un servidor o archivo y un cliente o servicio recolector de metadatos. El segundo puede pedir al primero que le envíe metadatos según determinados criterios. En respuesta el primero devuelve un conjunto de registros en formato XML, incluyendo identificadores (URL por ejemplo) de los objetos descritos en cada registro.

Las peticiones se emiten utilizando los métodos GET o POST del protocolo HTTP y constan de una lista de opciones con la forma de pares del tipo: clave=valor. Existen seis peticiones que un cliente puede realizar a un servidor: Para los proveedores de datos es importante tomar en consideración las siguientes definiciones y descripciones¹⁸:

¹⁷ Protocolo ampliamente utilizado para compartir catálogos de bibliotecas automatizadas. <http://www.bcl.jcyl.es/zeta/>

¹⁸ Traducción del OA-Forum Tutorial, <http://www.oaforum.org/tutorial/>

- Registros (*Records*)- Un registro son los metadatos de un recurso digital en un formato específico. Los registros están compuestos por tres partes: cabeza con metadatos (obligatorios) y una declaración descriptiva (opcional).
- Estampas de fecha (*Datestamps*)- Es la fecha de la última modificación de un registro y es obligatoria. La estampa permite ofrecer información acerca de los metadatos para búsquedas más selectivas utilizando desde (*from*) y hasta (*until*).
- Esquemas de metadatos (*Metadata schema*)- El OAI-PMH soporta la diseminación de múltiples formatos de metadatos de un repositorio, aunque todos deben de soportar como mínimo Dublin Core.
- Conjuntos (*Sets*)- Los conjuntos permiten la partición lógica de los repositorios y permite realizar búsquedas más selectivas (*set parameter*). No son obligatorias, no existen recomendaciones para su implementación y no son exhaustivas. Es importante y necesario que cada comunidad tenga acuerdos con respecto a la definición de sus conjuntos.
- Formato de petición (*Request format*)- Los repositorios deben soportar los métodos HTTP de GET y POST para peticiones. Existen seis tipos de peticiones: Identify, ListMetadataFormats, ListSets, ListIdentifiers, ListRecords, GetRecord. El cosechar de datos no necesariamente utiliza todas las peticiones pero un repositorio debe implementarlos todos.
- Respuesta (*Response*)- Las respuestas son tipo HTTP y el tipo de contenido debe de ser texto/xml.

Desde que fue aprobada la versión 2.0 de OAI-PMH en 2002, el protocolo ha permanecido estable y con mínimas modificaciones, su adopción generalizada debida a las pocas restricciones y la facilidad en su implementación posibilitó el acceso libre a los metadatos de numerosos repositorios en todo el mundo; hasta convertirse en una de las herramientas más utilizadas por las diferentes comunidades académicas para resguardar y hacer visible su producción.

4.3 Herramientas de software

En los últimos años se han desarrollado varios tipos de software que permiten crear y administrar un repositorio de cualquier tipo con relativa facilidad. Estas herramientas incorporan el protocolo de OAI-PMH como medio para permitir la interoperabilidad entre repositorios. Las soluciones técnicas más utilizados son CDSWare, ePrints y Dspace (Proberts Jenkins 2006). De acuerdo al servicio de ROAR (<http://archives.eprints.org/>), 160 repositorios utilizan Dspace y 209 Eprints. Dspace fue creado para capturar, describir y distribuir materiales digitales en formatos portadores de información de distintas y variadas naturalezas, mientras que ePrints está más orientado hacia repositorios de artículos académicos, manejo de documentos de texto principalmente (Proberts and Jenkins 2006).

A continuación se presenta una breve descripción de las soluciones técnicas propuestas en base a los resultados obtenidos en la fase de investigación de 3R. Para especificaciones técnicas, ver el Anexo A "Requerimientos técnicos de software"

4.3.1 CDSware

<http://cdsware.cern.ch/invenio/index.html>

A partir de julio de 2006 se llama CDS Invenio, es una aplicación desarrollada, utilizada y mantenida por el servidor de documentos del CERN, que es el laboratorio de física de partículas más grande del mundo y cuya producción intelectual arroja "documentos" en formatos digitales muy variados y en gran cantidad.

CDS Invenio es muy robusto y por lo mismo necesita muchos requerimientos de software y utilerías adicionales para su funcionamiento y explotación total de todas sus funciones. De acuerdo a su propia documentación esta diseñado para soportar repositorios muy grandes con materiales en diferentes formatos (puede manejar más de 350 colecciones de datos con más de 550,000 registros bibliográficos y 220,000 documentos a texto completo, aproximadamente).

Tiene funciones interesantes como extracción de referencias y búsqueda de palabras en el cuerpo del documento (en formatos MS Word, MS Excel, MS PowerPoint y PDF), presentación de resultados en múltiples formatos de salida (HTML, XML, MARC u OAI) y análisis de las bitácoras del log del servidor web, entre otras.

4.3.2 DSpace

<http://www.dspace.org/>

Es un software desarrollado por el MIT (Massachusetts Institute of Technology) con apoyo de Hewlet Packard (HP). Tiene muy en cuenta la visión de "comunidades"; es decir, una institución dedicada a la investigación está compuesta por departamentos, centros de investigación, participación de estudiantes y otras unidades y como cada una de estas unidades tiene requerimientos específicos, diferentes unos de otros, DSpace hace posible un flujo de trabajo y manejo de políticas que permite distribuir los contenidos, y manejar la propiedad intelectual de cada uno de ellos. También está orientado al problema de preservación a largo plazo de los materiales de investigación. Una característica importante de esta opción es el uso de PostgreSQL en lugar de MySQL como sistema manejador de bases de datos que es mucho más robusto y flexible además de que tiene un rendimiento excelente.

4.3.3 Fedora (Flexible Extensible Digital Object and Repository Architecture)

<http://www.fedora.info/>

Es un sistema de repositorios de propósito general desarrollado por la unidad de Ciencias de la Información de la Universidad de Cornell en conjunto con la Universidad de la Biblioteca de Virginia.

No es un software muy utilizado, pero de acuerdo con las características presentadas en la documentación es un software muy versátil: sus prerequisites de software son realmente mínimos y puede manejar eficientemente repositorios con hasta 1 millón de objetos en diferentes

formatos, se adapta fácilmente a otras aplicaciones o sistemas y comienza a tener un desarrollo importante en cuanto al manejo de políticas, manejador de versiones para los objetos almacenados y un mejor desempeño para soportar repositorios muy grandes.

Los usos más comunes que se le dan a este software son: gestión de colecciones bibliotecarias, repositorios de archivos, repositorios institucionales y bibliotecas digitales educativas.

4.3.4 Eprints

<http://www.eprints.org/>

Sus dependencias de software extra son mucho menores en comparación con CDS Invenio. De acuerdo con los propios desarrolladores, es un sistema en extremo fácil de poner en marcha y con un mínimo de experiencia técnica. Está pensado para ser muy flexible y adaptable a necesidades muy particulares.

Es un producto en constante desarrollo que comienza a incorporar funciones como búsquedas avanzadas, metadatos extendidos, etc. Una desventaja de su desarrollo continuo es que se está pensando cobrar por las características avanzadas aunque el núcleo del software y sus características funcionales básicas sigan siendo de libre distribución.

5. Interfaces de usuario en sistemas de búsqueda

Para el usuario final, un repositorio se visualiza como un sistema más de búsqueda de información. Las búsquedas representan una actividad compleja para muchos usuarios, por ello, dotarlos de interfaces adecuadas es un elemento que no debe dejarse de lado. La información más valiosa y el mejor sistema desarrollado pueden perder todas sus potencialidades sin la interfaz adecuada para el usuario. Por ello, para 3R, además de ser importante la tecnología que da vida al sistema del repositorio, también es importante la interfaz que comunicará al usuario con los objetos digitales que se le pueden proporcionar. Este apartado plantea la problemática del diseño de interfases en situaciones de búsqueda y recuperación de información.

5.1 El papel de la interfaz

La tecnología es el motor que impulsa el diseño de interfaces, permitiéndonos crear ambientes complejos a partir de técnicas de interacción y visualización cada vez más poderosas. Este motor es sin embargo un arma de dos filos: al tiempo que nos permite generar sistemas sofisticados y vistosos, nos ata con frecuencia a maneras de pensar que son contrarias a la dirección natural del comportamiento humano (Cooper, 1995). Integrar a los usuarios al desarrollo de software es un punto al que las Ciencias de la Computación le acuerdan cada vez mayor importancia. En efecto, existe hoy día una conciencia clara de que un software no sólo debe ser robusto, fiable, eficiente, conforme a estándares, etc.; sino que fundamentalmente debe satisfacer las necesidades del usuario, lo que se traduce en un conocimiento sobre su manera de hacer las cosas, su experiencia y sus preferencias personales.

La necesidad de este tipo de consideraciones se fundamenta en el rechazo que existe por parte de los usuarios ante aplicaciones que son funcionalmente correctas, pero con una lógica de uso difícilmente asequible. En estas situaciones, el rechazo puede ir desde frustración, el uso ineficiente del sistema, una muy alta tasa de errores, o un rechazo definitivo del sistema (Isis

Information Architects Inc., 2001; Scapin & Reynard, 1988).

En una época en que las computadoras se han convertido en un fenómeno de masas, es de vital importancia que la comunicación que se establece entre ellas y los usuarios sea lo más fluida y transparente posible. En otras palabras, hay que entender que para el usuario, la interfaz **es** el sistema.

Sin embargo, contrario a una creencia generalizada, el diseño de interfaces usuario no es un problema de sentido común, de estética (bonita y con colores agradables), o de capacitación. Se trata de una disciplina académica que reúne los métodos, modelos y criterios que permiten el estudio científico de las características del usuario: en primer lugar como ser humano (cuáles son sus capacidades y limitaciones para procesar información); en segundo lugar como trabajador (cuál es su tarea, cómo la percibe, qué objetivos persigue, qué informaciones y objetos manipula); y en tercer lugar, como usuario de un sistema interactivo (qué experiencia tiene en el manejo de computadoras, qué preferencias tiene, etc.)

Finalmente, la experiencia ha mostrado que la interfaz no es un módulo que se pueda generar en la fase final del proyecto, ya que su estructura y organización impacta de manera directa al resto de la aplicación. Es por ello que su diseño debe ser realizado desde las primeras etapas del ciclo de desarrollo y ser considerado como parte de la especificación, más que como el resultado de la estructura lógica de la futura aplicación (Gamboa, 1998).

5.2 El diseño de interfaces para la búsqueda de información

Vivimos en una sociedad en la que más que nunca la gente requiere de información mientras trabaja, aprende e incluso juega. El crecimiento en el volumen de la información disponible, así como en las necesidades particulares de los usuarios, ha llevado a que los modelos clásicos de búsqueda de información (ver Figura. 1), requieran de técnicas y herramientas innovadoras, fuertemente apoyadas por sistemas informáticos, que les permita satisfacer las demandas de los usuarios.

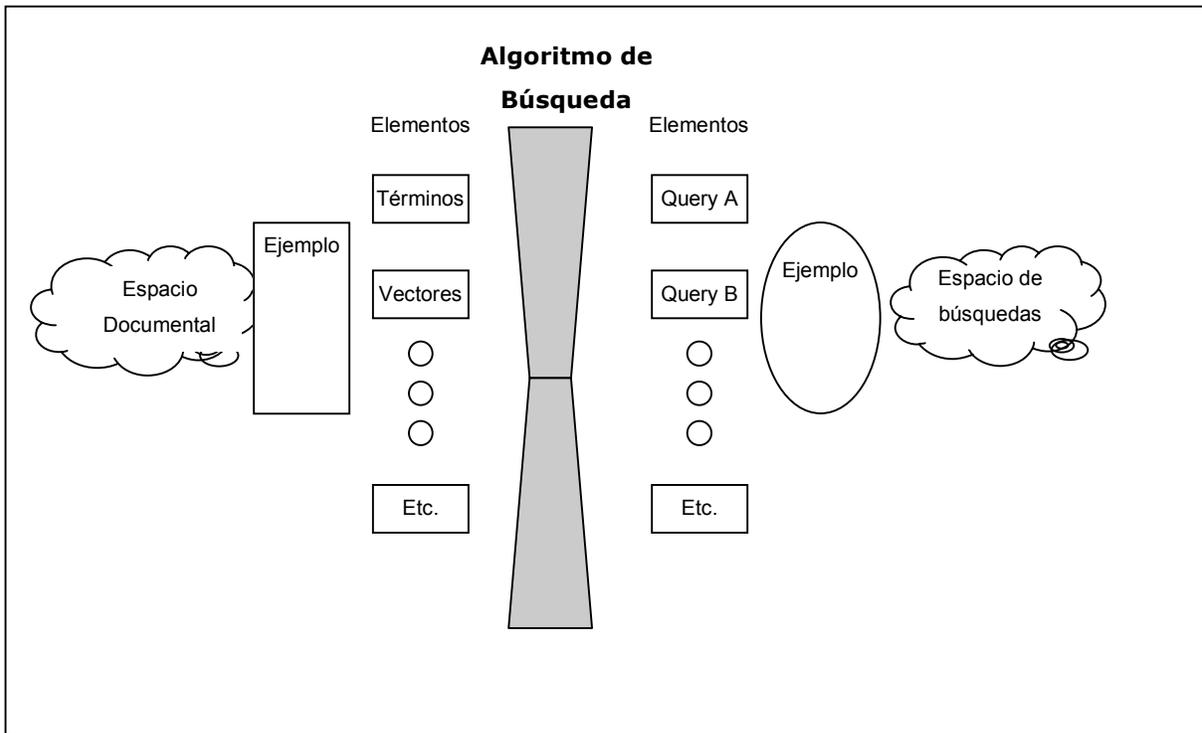


Figura. 1 Paradigma que ha dirigido la investigación en IR (Information Retrieval) por más de cincuenta años: un espacio de búsqueda que debe ser mapeado, a través de un conjunto de elementos y operadores, a un espacio documental. La métrica de evaluación es la exactitud con la que se empatan los dos espacios en términos de recuperación y precisión (Marchionini, 2004).

En este contexto, la búsqueda de información (IR de sus siglas en inglés), y la interacción humano computadora (HCI de sus siglas en inglés), se han convertido en dos áreas de investigación que con múltiples intereses y objetivos comunes: ambas estudian la manera en que los humanos procesan y organizan la información, concentran su atención, y toman decisiones. Cabe hacer notar que, como en otros casos, esta cercanía ha sido recientemente impulsada y potenciada por el gran desarrollo que han tenido las telecomunicaciones, en particular Internet.

Como ya se explicó, muchas de las coincidencias entre las áreas de IR y HCI surgen de su interés compartido sobre cómo el humano recibe, procesa, almacena, organiza y recupera información. Se considera en efecto que estos mecanismos cognitivos básicos son los que permiten al humano manipular

información, formarse una imagen mental del mundo exterior y tomar decisiones. Esto es de particular interés para ambas disciplinas, pues repercute de manera directa en la efectividad de los sistemas que pueden desarrollar, al involucrar aspectos como: cantidad de información que un humano puede manejar, tipo de asociaciones que puede realizar, maneras de estructurar y organizar la información, atención

5.3 El humano como procesador de información

Existe un enfoque alrededor de estos estudios en el que se defiende el "Modelo del Humano Como Procesador de Información" (Card, Moran, & Newell, 1983). Dicho modelo está fuertemente inspirado de la Inteligencia Artificial y de las arquitecturas de las computadoras modernas en el sentido que propone una estructura compuesta por un conjunto de memorias y de procesadores, cada uno dedicado a procesos específicos. El modelo, busca sintetizar y simplificar los procesos cognitivos hasta un punto en el que permite a los científicos modelar y predecir las reacciones de un humano frente a estímulos bien determinados en un contexto de trabajo mental. En el modelo, las memorias tienen características como: capacidad, duración y tiempo de codificación; mientras que los procesadores tienen características como: velocidad y funciones.

Si bien el modelo permite efectivamente analizar y predecir diversos comportamientos cognitivos en los humanos (tiempo que tarda en reaccionar, tiempo que tarda en actuar, cantidad de información que puede manejar, tiempo de retención de una información, etc.), la sobre-simplificación que demanda deja fuera otro gran número de fenómenos más complejos y de gran interés (toma de decisiones, formación de modelos mentales, etc.).

Existe por ello en el mundo académico otro enfoque acerca de cómo los humanos adquirimos información del mundo exterior y la utilizamos para tomar decisiones o resolver problemas. Esta visión se le conoce como "mecanismos distribuidos", en el sentido de que el humano requiere de diferentes niveles de información para poder formar una imagen del mundo

exterior y tomar una decisión.

5.4 El modelo cognitivo distribuido

Este enfoque es defendido por (Norman, 1990), quien establece que, en nuestra interacción con el mundo, los humanos debemos pasar por siete estados diferentes de procesamiento de información, iniciando con la elaboración de objetivos, generación de un plan, ejecución de acciones concretas, observación, análisis y evaluación y finalmente decisión.

Es el paso por cada uno de estos estados lo que nos permite ir formando un modelo mental del mundo y su funcionamiento. Eventualmente esto se traduce en experiencia y finalmente en posibilidad de predecir las consecuencias de una acción.

Como puede observarse, ambos enfoque son complementarios y ayudan a entender diferentes facetas del comportamiento mental humano. Ambos son muy utilizados tanto en IR como en HCI dado que buscan entender la manera en que los humanos manipulamos la información, y por esa vía buscan estar en posición de generar mejores y más eficientes herramientas que potencien la actividad mental de los usuarios.

5.5 El problema

En este proyecto consideramos que existe una oportunidad para mejorar de manera significativa las interfaces utilizadas en los sistemas de búsqueda de información. Como ya se mencionó, la interfaz ideal es la más simple posible, al tiempo que permite identificar fácilmente las funcionalidades principales del sistema.

Si bien los problemas que han enfrentado los buscadores de Internet en los últimos quince años han permitido avanzar en el problema de la interfaz de usuario, existen problemas no resueltos. Para Shneiderman (1997), estos problemas pueden observarse de manera indirecta en los resultados que aún obtenemos en una búsqueda: 1) listados interminables en los que el usuario debe posteriormente buscar la información que realmente está buscando; 2)

Cero resultados. En efecto, el autor señala que este tipo de resultados se debe en buena medida a que los usuarios no logran generar una idea clara sobre cómo funciona el sistema de búsqueda, y las implicaciones de sus acciones.

Existen diversos trabajos que estudian los problemas relacionados con la búsqueda de información en grandes bases de datos (Flynn, 1995; Hatlestad, 1996; Somerson, 1996). En ese sentido, existen también trabajos que muestran cómo una interfaz bien diseñada puede mejorar en mucho el desempeño y la satisfacción del usuario que utiliza el sistema (Koenemann & Belkin, 1996).

Nuestros trabajos se inscriben en la línea de las investigaciones hechas por (Rao et al., 1995; Van House, Butler, Ogle, & Schiff, 1996), en las que la especificación de la interfaz es precedida por un análisis de los usuarios y el contexto de uso en el que el sistema será utilizado. A estas consideraciones agregamos las reflexiones de (Marchionini, 2006), en el sentido que la información que actualmente el contenido ha ido más allá del simple texto al incluir imágenes, música, videos, código de cómputo, datos de sensores, secuencias bioquímicas, etc. La información es multimedia y multilingüe, y esta expansión en el tipo y formatos de datos sobrepasan los mecanismos de búsqueda e interacción clásicos. Aún más importante, el contenido se ha vuelto dinámico: por una parte el contenido cambia rápida y constantemente (como en los blogs, los wikis, etc.); pero además responde cada vez más a categorías difusas como las recomendaciones explícitas e implícitas de los usuarios, clasificaciones generadas por los usuarios, etc. Esto implica que si un algoritmo de indexado analiza un conjunto de documentos y genera una tabla de índice, los documentos ya habrán cambiado para el momento que el algoritmo termine. El punto es que los contenidos ahora tienen historia, y esta historia es importante para recuperar y eventualmente utilizar la información. Esto da pie a una nueva búsqueda de información basada en el contexto.

6. Análisis de Casos

Para ubicar el estado actual de los repositorios más reconocidos a nivel mundial y con la intención de identificar repositorios que puedan ser tomados como modelo para el caso del prototipo de 3R, se revisaron metódicamente repositorios a nivel internacional para describir el estado del arte en términos de software, recuperación de la información, organización y tipos de recursos electrónicos, políticas de ingreso y metadatos. El análisis de las interfases se ha dejado fuera de esta revisión dado que los repositorios estudiados manejan idiomas distintos y están enfocados a diferentes grupos de interés. En este apartado se presenta el análisis hecho a cada repositorio.

6.1 Metodología

Se revisaron de forma metódica los 704 repositorios en 44 países registrados en el índice de repositorios *Registry of Open Access Repositories* (ROAR), mencionado anteriormente. Se eligió este índice ya que cuenta con más registros que OpenDOAR, éste contaba con 379. El buscador OAIster incluye el mayor número de registros, 608, sin embargo, no es posible realizar búsquedas por país que eran pertinentes para nuestro estudio. La revisión se llevó a cabo durante el mes de junio del 2006.

De los repositorios revisados en ROAR se seleccionaron 145 ubicados en 24 países. La selección de estos repositorios se hizo en función de los siguientes criterios generales. Sin embargo, se encontraron algunos repositorios particularmente interesantes que se agregaron por otros criterios, en estos casos, se señala el criterio por el cual se realizó su inclusión a la lista final.

- El primer criterio es la tipología de los contenidos. En la UNAM nos enfrentaremos a distintos tipos de materiales, por lo tanto, los RI internacionales deben de contar con esa característica también. Se eliminaron los RI que sólo contienen *eprints* (pre o post) y se adoptó como criterio de selección aquellos que contaran una amplia variedad de

material tales como: artículos, conjuntos de datos de investigación, borradores, reportes, multimedia, presentaciones, disertaciones, objetos de aprendizaje, conferencias, capítulos, libros, software, tesis, proyectos, póster, revistas electrónicas. calendarios, programas, cursos, bases de datos, etc.

- El segundo criterio adoptado fue establecer en 1000 el número mínimo representativo de los registros seleccionados. Se prevé que el sistema prototipo de 3R tendrá que manejar grandes cantidades de información tomando en consideración el tamaño de la UNAM. Por lo tanto, se buscaron RI internacionales que manejaran un número mínimo de registros que se determinó en 1000. La gran mayoría de los RI tienen entre 50 y 800 registros y sobresalen algunos al tener más de 130,000 registros. Adicionalmente uno de los principales problemas de los RI es poblarlos y por lo tanto se eligieron casos exitosos de RI en donde se cuenta con un buen número de registros.
- El tercer criterio adoptado fue la fecha de actualización. Se estableció que la fecha del último ítem ingresado en el repositorio fuera posterior al año 2003.

6.2 Resultados

Los repositorios seleccionados se revisaron a detalle para obtener un panorama general del desarrollo y funcionamiento de los repositorios en el mundo académico, tener nociones básicas sobre su organización y base tecnológica.

En este primer rubro se observó si el repositorio seleccionado refleja la estructura de sus contenidos que lo integran y las comunidades los proveen; así como su integración y conexión con otras entidades. También se puso particular atención en aquellos RI que alcanzaron tasas elevadas de crecimiento en cortos periodos de tiempo.

Otro aspecto es el de la organización del tipo de material con que se organice el RI y se puso particular énfasis en los RI que manejan contenidos en diferentes formatos portadores de información (no solo texto sino también video, audio e

imágenes)

En la segunda parte se puso énfasis en el software más utilizado por los RI y sus capacidades respecto al manejo de distintos tipos de formatos digitales, gestión del número absoluto de registros y sus capacidades de búsqueda y formas de presentación de resultados.

6.2.1 Software

El software más utilizado en la implementación de RI a nivel mundial, según ROAR, son Eprints con 29,83%, Varios con 28,41% y DSpace con 22,44%. Ver Tabla 2.

Cuando nos referimos a la categoría de VARIOS nos referimos a desarrollos propios no exclusivos de software propietario que cumplen con el funcionamiento y estándares del protocolo OAI-PMH.

ROAR Software	Uso de software en 704 Repositorios ROAR 29/06/06	Porcentaje de uso en 704 Repositorios ROAR 29/06/06
E-Prints	210	29.82954545
Varios	200	28.40909091
DSpace	158	22.44318182
Bepress	50	7.102272727
ETD-db	23	3.267045455
OPUS	21	2.982954545
DiVA	14	1.988636364
CDSWare	9	1.278409091
ARNO	6	0.852272727
HAL	4	0.568181818
Fedora	3	0.426136364
DoKS	3	0.426136364
Edoc	2	0.284090909
MyCore	1	0.142045455
Total ROAR	704	100

Tabla 2: Porcentaje de uso de Software para RI a nivel mundial.

En el caso de los repositorios seleccionados 29,66% (43 de 145) usan DSpace, seguido por Varios con 26,21% (38 de 145) e Eprints con 17,24% (25 de 145). Ver Tabla 3.

Software	Uso de software en 145 repositorios elegidos	Porcentaje de uso en 145 repositorios elegidos
DSpace	43	29.66
Varios	38	26.21
E-Prints	25	17.24
Bepress	13	8.97
CDSWare	7	4.83
ARNO	6	4.14
HAL	6	4.14
DiVA	2	1.38
ETD-db	2	1.38
Fedora	1	0.69
OPUS	2	1.38
Total de Repositorios seleccionados	145	100.00

Tabla 3: Porcentaje de uso de Software para RI en una muestra de 145 repositorios

De los 145 repositorios seleccionados la distribución del software utilizado es como se muestra en la Figura 2.

También se realizó una cuantificación con del porcentaje relativo de software elegido respecto al total utilizado en los repositorios del ROAR (Figura 3), del total específico de cada software elegimos un porcentaje determinado; esto nos ayuda a determinar la calidad del software, si es que coincide en densidad con el número general de los repositorios que lo utilizan.

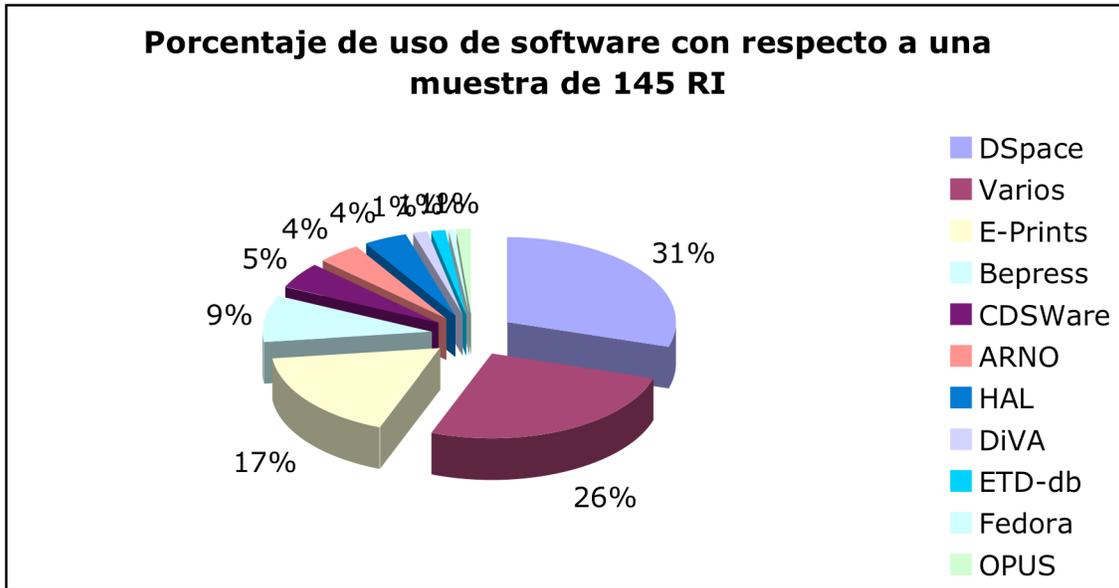


Figura 2: Distribución de uso de Software para RI en una muestra de 145 repositorios

ROAR Software	Número total en ROAR	Número seleccionado	Porcentaje seleccionado
E-Prints	210	25	11.9047619
Varios	200	38	19
DSpace	158	43	27.21518987
Bepress	50	13	26
ETD-db	23	2	8.695652174
OPUS	21	2	9.523809524
DiVA	14	2	14.28571429
CDSWare	9	7	77.77777778
ARNO	6	6	100
HAL	4	6	150
Fedora	3	1	33.33333333
DoKS	3	0	0
Edoc	2	0	0
MyCore	1	0	0

Tabla 4: Porcentaje de software elegido en relación con el número relativo de repositorios de ROAR que usan ese software

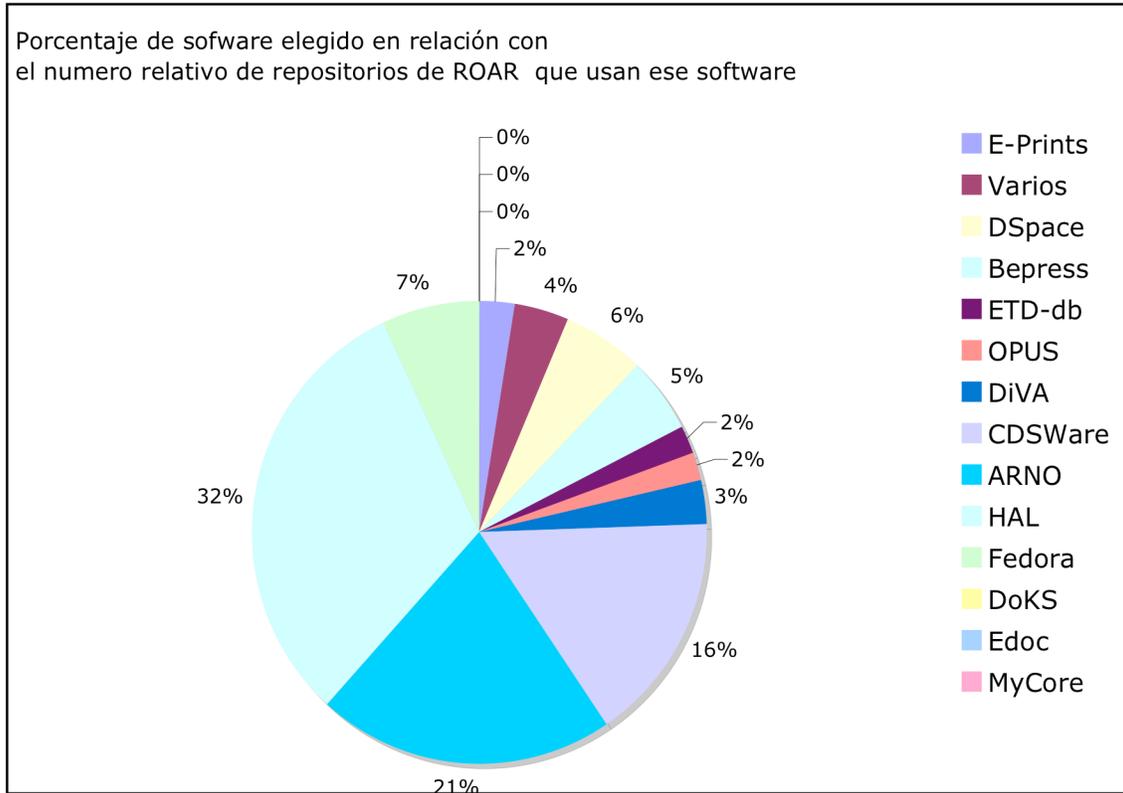


Figura 3: Porcentaje de software elegido en relación con el número relativo de repositorios de ROAR que usan ese software

6.2.2 Tipo de repositorio

Se detectaron tres tipos de repositorios en los registros:

- Repositorios que son en realidad revistas electrónicas. Este fue el caso en particular de Brasil. Se utiliza el software (Dpace o Eprints) para archivar los artículos y formar los números de las revistas. Estos no se tomaron en consideración para el estudio.
- Repositorios genéricos o temáticos- Estos son repositorios, generalmente desarrollados alrededor de un tema en particular o simplemente para que cualquier persona (generalmente después de registrarse) puede colocar material en el repositorio. Por ejemplo, E-LIS, *The open archive for*

*Library and Information Science*¹⁹, Estos no están desarrollados alrededor de una institución y no se han considerado pertinentes para las necesidades de esta fase del proyecto.

- Repositorios institucionales- Estos sistemas puestos en marcha por instituciones, generalmente universidades, para manejar el contenido académico de sus miembros. Sirve como escaparate, como una herramienta, para aumentar la visibilidad de lo que los académicos (y en ocasiones estudiantes de posgrado) están produciendo. Generalmente se organizan de acuerdo a la estructura académica (por Centros, Facultades, Institutos) y estos a su vez por temática. Estos repositorios son lo que más cercanamente reflejan las necesidades del proyecto 3R y fueron los seleccionados y para estudiarse a detalle.

El 100% de los 145 repositorios seleccionados son institucionales y reciben distintos tipos de materiales pero sólo 21 repositorios (13%) albergan contenidos diversos (no solo texto. Ver Figura 4).

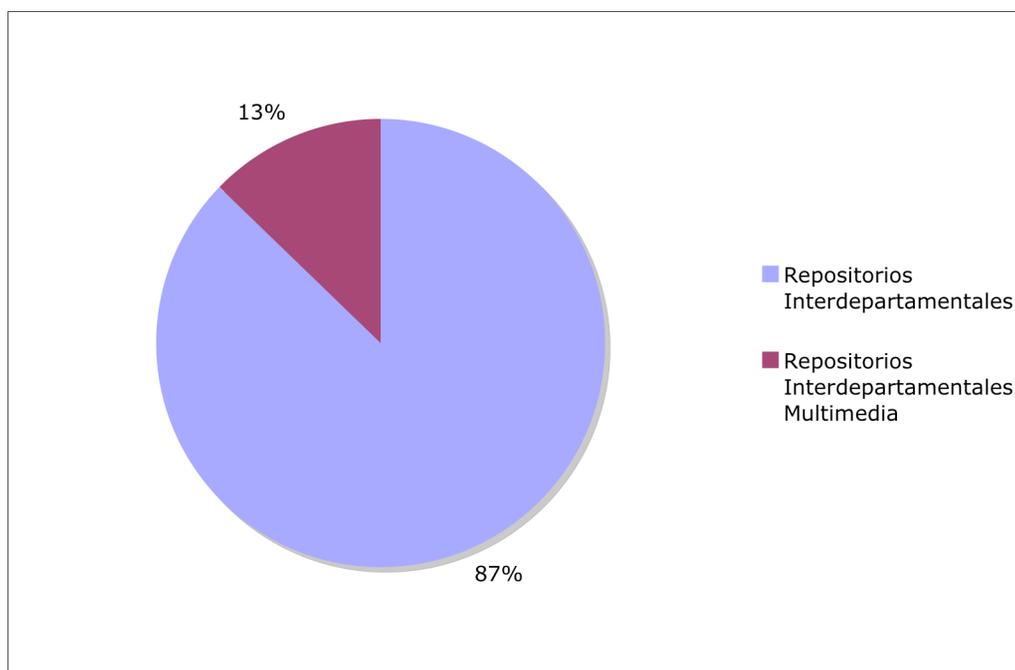


Figura 4: Relación entre los tipos de RI identificados

¹⁹ <http://eprints.rclis.org/>

Los repositorios seleccionados que entran en este 13%, es decir aquellos que aceptan formatos digitales diferentes a los usados comúnmente para texto (como es el caso de los formatos de audio y video) entre las soluciones de software que implementan están OPUS, CDSware y Eprints, que tienen capacidades de distribución de los objetos vía *streaming*; en otros casos como el de colecciones de imágenes y objetos de aprendizaje y video; DSpace y CDSware tienen la capacidad de realizar la agrupación de varios objetos ligados a un sólo registro.

6.2.3 Políticas de ingreso

En nuestro muestreo las políticas de ingreso de materiales a los repositorios varían según las disposiciones particulares y los objetivos de cada RI, de los 145 seleccionados el 95% requieren que los interesados en depositar materiales sean miembros de la institución académica, los demás permiten a usuarios externos el depósito. En otros, el 7%, es necesario el registro del usuario para la tener acceso completo a la consulta del material. En el caso de los repositorios que funcionan con el software Bepress, el 95%, es necesario pagar para tener acceso, de lo contrario el repositorio sólo permite visualizar los metadatos.

Estos datos deben ser afinados con una segunda visita detallada a la selección de repositorios

Existe un detalle que llama mucho la atención: la velocidad y la tasa de poblamiento; la respuesta puede estar en la estructura, las partes involucradas, en las políticas y mandatos institucionales, algunos repositorios presentan hasta 70 000 adiciones en un año, esta dinámica nos lleva a pensar en la problemática de las universidades de países con menos recursos cuyos repositorios esta poco poblados, sin mantenimiento, y con un sistema institucional sin mandatos ni políticas definidas.

7. Los repositorios en la UNAM

Dentro de la UNAM, la situación de los repositorios es incierta, no hay un registro o censo que pueda dar datos sobre cómo están este tipo de proyectos. Se sabe que hay esfuerzos importantes por desarrollar colecciones de recursos digitales, ya sea como bases de datos, bibliotecas digitales o depósitos de diversos tipos de contenidos. La tecnología utilizada es diversa y no hay políticas homogéneas para el ingreso y manejo de los recursos.

Después de la investigación realizada en esta fase del proyecto de 3R, en este apartado se presenta, partiendo de las definiciones previamente dadas sobre los distintos tipos de repositorio, una primera propuesta sobre lo que se puede considerar un repositorio universitario. Así mismo se hace un intento por comenzar a tipificar los diferentes tipos de repositorios en la Universidad y se presenta el análisis de los primeros repositorios candidatos para integrarse al prototipo de 3R.

7.1 Los repositorios universitarios

La UNAM tiene características y necesidades propias en cuanto a la producción y gestión de los recursos digitales que genera, así que derivado del análisis de las diversas definiciones de repositorios que se encuentran a nivel mundial y considerando las particularidades de la comunidad universitaria, a continuación se describe lo que se considera propio de un repositorio universitario (RU).

Los repositorios universitarios que se integren a 3R:

- Contendrán materiales digitales producto intelectual de por lo menos un miembro académico universitario. Los miembros académicos de la UNAM son los profesores, investigadores y técnicos académicos. Posteriormente, se podrá contemplar la opción de incluir estudiantes del posgrado
- Contendrán el recurso mismo y no solamente sus metadatos
- Contarán con un responsable académico y un responsable técnico

avalados por la institución que implementa el repositorio.

- Deberán implementar el protocolo OAI-PMH para la interoperabilidad con otros repositorios universitarios. Esto incluye el estándar mínimo de metadatos de Dublin Core
- Utilizarán un software de aplicación específica, recomendado por 3R o compatible, para la creación y el manejo del repositorio

Cada repositorio universitario definirá a nivel local:

- El tipo y formato de materiales ingresados
- La asignación de permisos para depositar y asignar metadatos
- El software para la creación y administración del repositorio a partir de los recomendados por 3R o alguno compatible
- Esquemas de metadatos adicionales al Dublin Core
- Otras políticas generales de administración de los materiales

Estas definiciones son un primer acercamiento a lo se busca como repositorio universitario, conforme el proyecto se vaya desarrollando la definición puede modificarse y seguramente irá consolidándose.

Por ahora, con estos criterios generales se podrán evaluar los actuales repositorios de la UNAM, candidatos para incorporarse a 3R y así definir cuáles podrían utilizarse como parte del prototipo. Los repositorios que cumplan con estas características se denominarán repositorios universitarios.

7.2 Selección de repositorios candidatos

En esta etapa del proyecto se buscó detectar los repositorios candidatos para el desarrollo del prototipo de 3R. Uno de los principales problemas que se busca resolver con la creación de una red de repositorios universitarios, es la detección y visibilidad de los recursos universitarios. Por lo tanto, encontrar los repositorios existentes para la creación de la red por consecuencia no es una tarea sin sus complicaciones. Andrew (2003) en un estudio para evaluar los recursos electrónicos colocados en repositorios por investigadores, concluye que una evaluación sistemática es preferible a censos por cuestionarios o muestreo. Por lo tanto, se partió de la página principal o raíz de la institución

(en este caso www.unam.mx) y se siguieron los vínculos hacia las páginas de segundo y tercer nivel, con el objetivo de detectar los repositorios existentes.

Los resultados indican que la estructura del dominio unam.mx es altamente compleja y profunda. Encontramos aproximadamente 2,170,000 de páginas dentro del dominio registrados por Google. Yahoo! registra 2,084,788. Esto dificulta el proceso sistemático debido al volumen.

Así mismo, no existe una forma normalizada de elaborar los URL de acuerdo a las distintas dependencias. Esto es el caso tanto en la nomenclatura como en su relación jerárquica administrativa. La Tabla 5 contiene ejemplos de URL y las dependencias a las que pertenecen poniendo en claro la dificultad para encontrar RI al interior de la Universidad siguiendo esta metodología.

Dependencia	URL
Instituto de Investigaciones Antropológicas	http://swadesh.unam.mx/home.htm
Instituto de Investigaciones Filológicas	http://www.filologicas.unam.mx/
Difusión Cultural UNAM	http://difusion.cultural.unam.mx/
Facultad de Economía	http://herzog.economia.unam.mx/
Centro de Ciencias Físicas de la UNAM	http://www.fis.unam.mx/
Poesía de José Juan Tablada	http://www.tablada.unam.mx

Tabla 5: Ejemplos de URL encontrados en la UNAM

Por lo mencionado, el análisis de navegación metódica no presentó los resultados esperados y no se detectaron repositorios universitarios; de manera complementaria se solicitó al comité de 3R, cuyos miembros cuentan con experiencia en la creación de recursos electrónicos, su asesoría para nombrar algunos repositorios candidatos, centrándose en algunos los candidatos de la propuesta original (Castro et. al, 2006).

Se sostuvieron reuniones con los encargados de los repositorios candidatos y se analizaron las características principales de cada propuesta. Posteriormente se analizó su propuesta a partir de los criterios determinados por nosotros de

un repositorio universitario y se evaluó la factibilidad de que participaran en el prototipo. A continuación se presenta el análisis por repositorio.

7.3 Instituto de Biología- UNIBIO

La Unidad de Informática para la Biodiversidad y el Ambiente forma parte del Instituto de Biología y se encargan del desarrollo de herramientas tecnológicas para apoyar la investigación. La UNIBIO actualmente se encuentra desarrollando un repositorio institucional y existe la disposición para convertirlo en un repositorio universitario para la red de 3R. En un principio el repositorio contendrá la información de sus colecciones biológicas que contienen trabajo de sus investigadores en torno a la clasificación y descripción de las especies, acompañado por imágenes de las mismas. Posteriormente se planea incluir todo tipo de materiales, incluyendo los artículos en forma de eprints, relacionados con las colecciones. Así mismo, existen mapas georeferenciados, fotocolectas, bases de datos, presentaciones de conferencias y otros materiales producidos en el Instituto.

La UNIBIO cuenta con la experiencia técnica y están dispuestos a instalar y mantener su propio servidor. Así mismo cuentan con experiencia en catalogación y clasificación. Su objetivo es añadir metadatos adicionales al Dublin Core (en particular el Darwin Core) y este trabajo será útil para ver como se añaden esquemas.

7.4 BIDI-MADI

La Biblioteca digital de Materiales Didácticos, pertenece al macroproyecto de Tecnologías de Información, etc. y está formado por integrantes de varias dependencias de la UNAM. Su objetivo es detectar materiales didácticos, tanto en formato análogo como en digital, y catalogarlos con el principal objetivo de poner a disposición de los universitarios la información acerca de los materiales didácticos que han sido creados dentro de la UNAM.

Debido a que el proyecto BIDI-MADI contempla todo tipo de materiales, en el

caso de 3R sólo serían pertinentes materiales digitales en línea. Unos de los retos sería detectar y extraer solamente los adecuados. En este sentido la experiencia podría servir ya que probablemente este será una forma de poblar en el futuro los repositorios, extrayendo de catálogos existentes.

Sin embargo, BIDI-MADI está concebido como una colección temática, y no institucional. En este sentido los repositorios universitarios tienen un enfoque institucional más que temático. El prototipo de BIDIMADI está centrado en recolectar los materiales didácticos de 12 dependencias universitarias. Sería necesario para su inclusión a la red de repositorios, seleccionar solamente una de estas dependencias y que el repositorio fuera esa institución, aunque para la construcción del prototipo el contacto fuera BIDI-MADI. En caso de ser exitoso se podrían añadir posteriormente las once dependencias restantes como repositorios universitarios. En este sentido esta experiencia nos permitiría escalar rápidamente.

7.5 BIDI-UNAM

La Biblioteca Digital de la UNAM reúne más de 70,000 recursos electrónicos y claramente es la colección de recursos digitales más grande de la universidad. La BIDI pertenece a la Dirección General de Bibliotecas. BIDI-UNAM da acceso a numerosos recursos gratuitos así como a los materiales suscritos, principalmente revistas electrónicas, bases de datos y libros.

La BIDI-UNAM, sin embargo no es un repositorio sino una biblioteca digital y ofrece principalmente materiales producidos por autores externos a la UNAM. Al igual que BIDI-MADI, BIDI-UNAM tiene un enfoque suprainstitucional.

Sin embargo, existe una sección de la BIDI-UNAM que se ha enfocado a recavar recursos electrónicos producidos en la UNAM. Se espera poder recuperar esos materiales y proponer a la institución de los creadores, actuar como repositorio universitarios. Así mismo, la BIDI-UNAM cuenta con el software de HERMES y actualmente está trabajando con OAI-PMH y será importante apoyarse en esa experiencia para la creación de la red.

7.6 Instituto de Investigaciones Sociales (IIS)

En la entrevista realizada se manifestó, por parte del ISS, el interés y disponibilidad para contar con un mecanismo de difusión del trabajo de investigación producido por el centro; a la fecha el instituto cuenta con una planta de 70 investigadores y ocho áreas generales de investigación, muchos de sus proyectos están ligados a otros de alcance nacional.

El punto de partida serán los contenidos relacionados con el tema de las políticas públicas, los movimientos sociales y su relación con la lucha por los recursos ambientales, esta área temática proporcionaría dentro del marco del prototipo 3R, posibilidades para establecer vínculos con las bases de datos sobre medio ambiente del Instituto de Biología- UNIBIO, por lo tanto, este repositorio desde la perspectiva de la investigación social y las humanidades tiene posibilidades idóneas de vinculación e integración a otras experiencias del macroproyecto.

Actualmente los investigadores del IIS participan con la publicación de sus trabajos en redes de intercambio académico, algunos depositan sus artículos en repositorios temáticos y bibliotecas digitales como CLACSO. Este antecedente sugiere que podrán utilizar con entusiasmo y agilidad las posibilidades del repositorio.

La propuesta del repositorio universitario del IIS incluiría todo tipo de material: bases de datos, una de ellas con 2000 registros, *eprints*, fotografías, presentaciones en congresos, libros, tesis doctorales, materiales de audio y video, entre otras.

El IIS cuenta con un área técnica de apoyo en cómputo, así como el hardware necesario para crear el repositorio. Su participación en el proyecto consistiría en recibir los lineamientos básicos de 3R e implementarlos en un contexto institucional, ampliar sus aspectos temáticos y contribuir a la elaboración de políticas y mejores prácticas para otros institutos de humanidades.

La variedad y riqueza del material que puede proporcionar este instituto, el

entusiasmo y el apoyo institucional (Secretaría Académica del instituto hizo una solicitud de participación en el macroproyecto) hacen de este repositorio un candidato ideal para su incorporación al prototipo 3R.

7.7 Colecciones mexicanas – DGSCA

Es una base de datos en línea que se compone de tres fondos, Archivo Franciscano, Revistas Literarias del Siglo XIX y Españoles en México en el Siglo XIX. Este acervo fue implementado por DGSCA en 2003, inicialmente con este desarrollo se probó el software de repositorios DSpace para sus tres colecciones; la posible elección de este acervo como candidato a 3R estará determinada por la propiedad y origen de los materiales, ya que las dos primeros acervos no son de autoría del personal académico de la UNAM, son solo recopilaciones y digitalizaciones de materiales que no pertenecen a la Universidad; el tercero, Españoles en México en el Siglo XIX, fue compilado en el marco de un proyecto de investigación del Instituto de Investigaciones Bibliográficas (IIB), por el Dr. Pablo Mora, el material pertenece a la Biblioteca Nacional de México, este es un recurso que por su procedencia cumple con el criterio institucional requerido por el prototipo 3R. Colecciones mexicanas presenta también como desventajas el hecho de que las colecciones se tendrían que dividir y al ser el fondo Españoles en México parte de la investigación del IIB, el repositorio debería ser implementado y gestionado por este instituto, ambas circunstancias descartan la elección de este repositorio para esta etapa del proyecto.

7.8 Centro de Investigaciones en Energía (CIE)

El Centro de Investigaciones en Energía de la UNAM, localizado en Temixco, cuenta actualmente con 56 investigadores y 75 estudiantes de posgrado. Los temas principales de investigación son materiales solares, sistemas energéticos y termociencias. Se sostuvieron dos reuniones con un investigador y miembros de centro de cómputo. La propuesta del repositorio universitario del CIE incluiría todo tipo de material, incluyendo *eprints*, simulaciones en java,

software, tesis doctorales entre otras.

Se considera que los investigadores del CIE, principalmente los físicos, conocen el sistema de repositorios y de *eprints*. Algunos de ellos ya depositan sus artículos en repositorios temáticos tales como el arXiv. Esto facilita la creación del repositorio y sugiere que habrá entusiasmo y poca resistencia. El CIE cuenta con el apoyo técnico, así como el hardware necesario para crear el repositorio universitario. Su participación en el proyecto consistiría en recibir los lineamientos básicos de 3R e implementarlos en un contexto institucional. Así mismo, se encargarán de detectar las necesidades de un instituto como este y contribuir a la elaboración de políticas y mejores prácticas para otros institutos.

El CIE ya participa en el macroproyecto de Tecnologías de la Información, en el Observatorio Técnico. Se considera, por lo tanto, que tiene posibilidades idóneas de vinculación e integración a otras experiencias del macroproyecto.

Debido a que el CIE se encuentra fuera del campus universitario, podrá servir como experiencia para evaluar las mejores prácticas de trabajo con organismos universitarios externos. Esta experiencia será particularmente útil para que el prototipo incluya consideraciones para la creación de repositorios universitarios geográficamente dispersos.

Actualmente el CIE no cuenta con ningún tipo de repositorio centralizado. El material electrónico se encuentra en sus páginas personales y en repositorios temáticos, así como fuera de línea en máquinas locales. Sin embargo, el perfil del personal permite pensar que el CIE puede funcionar como *early adopters*. Kling (2003) notó las diferentes actitudes disciplinarias hacia los recursos electrónicos, así como su deseo de compartirlas. Particularmente en la comunidad de físicos (Harnard 2001) se ha detectado este fenómeno. El CIE ofrecerá material creado por universitarios, estarán disponibles en línea y formato digital, se apegarán a las normas básicas de interoperabilidad y utilizarán el Dublin Core.

8. Reflexiones finales y trabajo futuro

Los repositorios de contenidos digitales están en fase de crecimiento y desarrollo, por lo cual, no hay un esquema único que determine una concepción inequívoca. Esto permite que cada organización adecúe el concepto a su propias necesidades y objetivos.

Aunque no hay un único tipo de repositorios, se marca una tendencia generalizada en búsqueda de la visibilidad de contenidos y de la interoperabilidad de sistemas, esto conlleva la utilización de tecnologías estándares principalmente orientadas a su operación en Internet, que permiten comunicación entre diversos repositorios y facilitan los procesos de búsqueda en el universo de los contenidos digitales. Hacer uso de estas tecnologías facilita al escalabilidad y evolución de los repositorios desarrollados, así como también hacen más viable compartir recursos entre sistemas diversos, aún más entre repositorios similares o de una misma organización.

Las herramientas de software se perciben bien logradas, con funcionalidades mínimas que permiten crear repositorios funcionales y estables, lo que aligera el camino de aquellos que quieren emprender un proyecto de repositorio. Se puede adelantar que la visión del proyecto Sherpa es totalmente compartida con lo que se busca llevar a acabo con la implementación 3R, por lo que para futuras fases del proyecto se tomará especial cuidado en la revisión de CDS Invenio, así como se ve también sugerente el uso de DSpace, debido a que este software permite la manipulación de objetos en múltiples formatos digitales.

El análisis de casos llevó a identificar que hay diversidad en cuanto al tipo de repositorios y que cada organización adopta (y adapta) el que mejor convenga a sus propios objetivos, aunque para el estudio se tomaron sólo los repositorios institucionales, por considerarse lo más pertinentes para esta etapa del proyecto. Se identificó que los contenidos son diversos en contenido y forma, lo cual abre el espectro de cobertura de un repositorio contra el de

una biblioteca digital. El análisis también permitió ver que el software más utilizado está entre DSpace y E-prints, uno abierto a diversos tipos de contenidos y otro específico para textos. Así mismo, se observó que el acceso a los contenidos y a los metadatos es variable, algunos permiten consulta abierta de todos, algunos otros sólo a través de registro y otros incluso requieren pago. En su gran mayoría, un repositorio sólo recibe contribuciones de miembros de la organización.

En la UNAM, no se encuentra un desarrollo formal de repositorios, de hecho no se pudo encontrar un repositorio estricto, formalmente definido y operable como repositorio. Los proyectos que se estudiaron se ubican más dentro de los ámbitos de colecciones o de bibliotecas digitales, aunque con algunas adaptaciones podrán fácilmente convertirse a repositorios universitarios o al menos podrán mantener comunicación con éstos. Afortunadamente, se encontraron proyectos interesantes que van dando inicio y que están dispuestos a seguir las políticas que puedan desarrollarse para el crecimiento de los repositorios universitarios.

Aunque se encontró mucha información valiosa en este trabajo, esta primera etapa de investigación no es concluyente y se irá dando más consistencia al tema de los repositorios universitarios, y a la red misma, con las siguientes etapas de este proyecto.

Los siguientes pasos estarán encaminados a la aplicación de la teoría y de los proyectos que se evaluaron para poder determinar un modelo conceptual que refleje la tecnología, la operación y las políticas que 3R deba seguir, al menos en el modelo prototipo.

Para la validación de las siguientes etapas se está invitando a expertos de diversos ámbitos (expertos en Internet, programación, repositorios, contenidos, entre otros) para que evalúen la propuesta y enriquezcan lo que este grupo de trabajo pueda proponer.

9. Referencias

- Andrew, T. (2003). "Trends in Self-Posting of Research Material Online by Academic Staff." *Ariadne* [Revista electrónica], 37
<http://www.ariadne.ac.uk/issue37/andrew/>
- Atkins, D. E. C., K. K. Droegemeier, et al. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation - Blue Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation, 84.
http://scholar.google.com/url?sa=U&q=http://www.cs.odu.edu/~keyes/scales/reports/cyberinfra_2003.pdf
- Awre, Chris, Hanganu, Gabriel, Ingram, Carolin, Brett, Tony, Dolphin, Ian (2005). "The CREE Project. Investigating User Requirements for Searching within Institutional Environments", *D-Lib Magazine* [Revista electrónica] 11 (10).
<http://www.dlib.org/dlib/october05/awre/10awre.html>
- Barton Mary R., and Margaret M. Waters (2004-2005). *LEarning About Digital Institutional Repositories, Creating an Institutional Repository: LEADIRS Workbook*, The Cambridge-MIT Institute (CMI)
<http://www.dspace.org/implement/leadirs.pdf>
- Barrón Ruiz, Angela, et al (2006). *Propuesta de evaluación de objetos de aprendizaje* [Presentación Microsoft Power Point] Universidad de Salamanca.
<http://www.uoc.edu/symposia/spdece05/ppt/ID06.ppt>
- Barrueco José Manuel, Imma Subirats (2003). "Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo" *El Profesional de la Información*, 12 (2), 99-106,
<http://eprints.rclis.org/archive/00000177/02/cardedeu.pdf>
- Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago, American Library Association.
- Card, S., Moran, T. P., & Newell, A. (1983). *The Psychology of Human Computer Interaction*. Prentice Hall.

- Castro, Alberto, Gamboa, Fernando, Galina, Isabel, Gímenez, Joaquín, López G., Clara, (2006)
Red de Repositorios Universitarios de Recursos Digitales. Propuesta de proyecto de
Investigación para el Macroproyecto Tecnologías para la Universidad de la Información y
la Computación , UNAM, México.
<http://eprints.rclis.org/archive/00006324/>
- Chan Leslie (2004). "Supporting and Enhancing Scholarship in the Digital Age: The Role of Open-
Access Institutional Repositories", *Canadian Journal of Communication* [Revista
electrónica], 29 (3).
http://eprints.rclis.org/archive/00002590/01/Chan_CJC_IR.pdf
- Cooper, A. (1995). *About face. The essentials of user interface design* (1rst ed.). New York, NY:
IDG Books Worldwide.
- Crow Raym (2002). *The Case for Institutional Repositories: A SPARC Position Paper*.
The Scholarly Publishing & Academic Resources Coalition (c) SPARC
<http://www.arl.org/sparc/IR/ir.htm>
- Flynn, L. (1995, 2 October). "Making searches easier in the web's sea of data". New York Times.
- Gamboa, F. (1998). *ALACIE: Atelier Logiciel d'Aide à la Conception d'Interfaces Ergonomiques*.
Unpublished Thèse de Doctorat, Paris 11, Paris.
- Harnard, S. (2001). "The Self-Archiving Initiative." *Nature* 410. 1024-1025.
<http://www.ecs.soton.ac.uk/~harnad/Tp/nature4.htm>
- Hatlestad, L. (1996, 23 september). "Internet search not over yet". Infoworld, 18.
- Heery, Rachel, Anderson, Sheila (2005). *Digital Repositories Review UKOLN*, University of Bath
Arts and Humanities Data Service.
http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf
- Hey, T. and A. E. Trefethen (2005). "Cyberinfrastructure for e-Science."
Science 308(5723), 817-821.
<http://eprints.ecs.soton.ac.uk/10844/>
- Isis Information Architects Inc. (2001). *Interface Hall of Shame*, Disponible en:
<http://www.iarchitect.com/mshame.htm>
- Johnson, Richard K. (2002). "Institutional repositories: partnering with faculty to enhance
scholarly communication"; *D-Lib Magazine*, November 2002.
<http://www.dlib.org/dlib/november02/johnson/11johnson.html>

- Jones, R., Theo A. and MacColl, J. (2006), *The Institutional Repository*, Oxford. 250 pp.
Review by Sally Rumsey at *Ariadne*, Issue 47 April 2006
<http://www.ariadne.ac.uk/issue47/rumsey-rvw/>
- Kling, Rob. and McKim, Geoffrey (1999). "Scholarly Communication and the Continuum of Electronic Publishing." *Journal of the American Society for Information Science* 50(10), 890-906.
<http://www.citebase.org/fulltext?format=application%2Fpdf&identifier=oai%3AarXiv.org%3Acs%2F9903015>
- Koenemann, J., & Belkin, N. (1996). "A case for interaction: A study of interactive information retrieval behavior and effectiveness". In ACM (Ed.), *CHI 96, Human Factors in Computing Systems* (pp. 205-212). New York, N.Y.: ACM Press.
- Lawrence, S. (2001). "Free online availability substantially increases a paper's impact" *Nature*, 31
<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- López, Clara. (2005) *Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning*, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo)
- Marchionini, G. (2006). *Toward Human-Computer Information Retrieval*. Retrieved Access Date, Access 2006
- Norman, D. A. (1990). "The design of everyday things". New York: Doubleday.
- Proberts, Steve y Jenkins, Celia (2006). "Documentation for Institutional Repositories", *Learned Publishing*, 19 (1).
<http://magpie.lboro.ac.uk/dspace/bitstream/2134/782/1/lppaper+-+final.pdf>
- Scapin, D. L., & Reynard, P. (1988). *La conception ergonomique des interfaces: problèmes de méthode* (Rapport Technique No. 952). Rocquencourt, France: Institut National de Recherche en Informatique et en Automatique.
- Shneiderman, B., Byrd, D., & Croft, W. B. (Enero, 1997). *Clarifying Search, A User-Interface Framework for Text Searches*. D-Lib Magazine.
- Somerson, P. (1996, August 97). Web Coma. PC Computing.
- Suber, P. (2004). Open Access Overview.
<http://www.earlham.edu/~peters/fos/overview.htm>

Unsworth, J. C., P. Courant, et al. (2005)

. *The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences* (for public comment), 64.

Van House, N., Butler, M., Ogle, V., & Schiff, L. (1996). *User-Centered Iterative Design for Digital Libraries: The Cypress Experience*. Retrieved Access Date, Access 1996,
Disponible en <http://www.dlib.org/dlib/february96/02vanhouse.html>

10. Bibliografía recomendada

- (2006) " Access to Literature: The debate continues." *Nature*, 2006.
<http://www.nature.com/nature/focus/accessdebate/index.html>
- Andrew, T. (2003). "Trends in Self-Posting of Research Material Online by Academic Staff." *Ariadne* [Revista electrónica], 37
<http://www.ariadne.ac.uk/issue37/andrew/>
- (2005) "Announcing Symposia: New institutional repository system from Innovative Interfaces". *Presswire* [Revista electrónica], 4
http://www.iii.com/news/pr_template.php?id=238
- Aparac, T. (Ed.), *Digital libraries: Interdisciplinary concepts, challenges and opportunities*, pp. 29-42. Zagreb: Benja.
http://www.ffzg.hr/infoz/colis3/dlib_perspectives.ppt
- Archives, O. (2004). Open Archives Initiative- Frequently asked questions (FAQ).
<http://www.openarchives.org/documents/FAQ.html>
- Atkins, D. E. C., K. K. Droege, et al. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation - Blue Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation, 84.
http://scholar.google.com/url?sa=U&q=http://www.cs.odu.edu/~keyes/scales/reports/cyberinfra_2003.pdf
- Awre, Chris, Hanganu, Gabriel, Ingram, Carolin, Brett, Tony, Dolphin, Ian (2005). "The CREE Project. Investigating User Requirements for Searching within Institutional Environments", *D-Lib Magazine* [Revista electrónica] 11 (10).
<http://www.dlib.org/dlib/october05/awre/10awre.html>
- Barrón Ruiz, Angela, et al (2006). *Propuesta de evaluación de objetos de aprendizaje* [Presentación Microsoft Power Point] Universidad de Salamanca.
<http://www.uoc.edu/symposia/spdece05/ppt/ID06.ppt>
- Barrueco José Manuel, Imma Subirats (2003). "Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH): descripción, funciones y aplicación de un protocolo" *El Profesional de la Información*, 12 (2), 99-106,
<http://eprints.rclis.org/archive/00000177/02/cardedeu.pdf>
- Barton, Mary R. and J. Harford (2003). "Building a Business Plan for DSpace, MIT Libraries Digital Institutional Repository." *Journal of Digital Information* [Revista electrónica], 4(2f).
<http://jodi.tamu.edu/Articles/v04/i02/Barton/>
- Barton Mary R., and Margaret M. Waters (2004-2005). *LEarning About Digital Institutional Repositories, Creating an Institutional Repository: LEADIRS Workbook*, The Cambridge-MIT Institute (CMI)
<http://www.dspace.org/implement/leadirs.pdf>
- Budapest Open Access Initiative. *A Guide to Institutional Repository Software v 3.0*
<http://www.soros.org/openaccess/software/>
- Budapest Open Access Initiative, *Guide to Open Access Publishing and Scholarly Societies*.
http://www.soros.org/openaccess/pdf/open_access_publishing_and_scholarly_societies.pdf
- Caplan, P. (2003). *Metadata Fundamentals for All Librarians*. Chicago, American Library Association.
- Carr, L. and Harnad, S. (2005) *Keystroke Economy: A Study of the Time and Effort Involved in Self-Archiving*
<http://eprints.ecs.soton.ac.uk/10688/01/KeystrokeCosting-publicdraft1.pdf>
- Castro, Alberto, Gamboa, Fernando, Galina, Isabel, Gímenez, Joaquín, López G., Clara, (2006) Red de Repositorios Universitarios de Recursos Digitales. Propuesta de proyecto de Investigación para el Macroproyecto Tecnologías para la Universidad de la Información y la Computación, UNAM, México.

- <http://eprints.rclis.org/archive/00006324/>
Chan, Leslie and B. Kirsop (2001). "Open Archiving Opportunities for Developing Countries: towards equitable distribution of global knowledge." *Ariadne* [Revista electrónica], 30.
<http://www.ariadne.ac.uk/issue30/oai-chan/>
- Chan Leslie (2004). "Supporting and Enhancing Scholarship in the Digital Age: The Role of Open-Access Institutional Repositories", *Canadian Journal of Communication* [Revista electrónica], 29 (3)
http://eprints.rclis.org/archive/00002590/01/Chan_CJC_IR.pdf
- (2006). "Colby College selects Symposia to build institutional repository." *Presswire* [Revista electrónica], 22
http://www.iii.com/news/pr_template.php?id=275
- Coleman Anita, Paul Bracke, S. Karthik (2004). "Integration of Non-OAI Resources for Federated Searching in DLIST, an Eprints Repository", *D-Lib Magazine* [Revista electrónica] 10 (7/8)
<http://www.dlib.org/dlib/july04/coleman/07coleman.html>
- Committee, J. I. S. (2006). "Focus on Access to Institutional Repositories", *Joint Information Systems Committee*, 28.
http://www.jisc.ac.uk/uploaded_documents/JISC-FAIR-Synthesis_FINAL.pdf
- Crow Raym (2002). *The Case for Institutional Repositories: A SPARC Position Paper*. The Scholarly Publishing & Academic Resources Coalition (c) SPARC
<http://www.arl.org/sparc/IR/ir.htm>
- "Dspace. An Open Source Dynamic Digital Repository"
D-Lib Magazine, January 2003. Volume 9 Number 1
<http://www.dlib.org/dlib/january03/smith/01smith.html>
- Jacinto A. Dávila, Luis A. Núñez, Beatriz Sandia Y Rodrigo Torrén (2005). *Los repositorios institucionales y la preservación del patrimonio intelectual académico*, Borrador, Merida, Ven.
<http://webdelprofesor.ula.ve/ciencias/nunez/Articulos/Teleinformacion/AccesoLibreConciamientoInterciencia050809.pdf>
- Jacinto A. Dávila, Luis A. Núñez, Beatriz Sandia, José G. Silva Y Rodrigo Torrén (2006) *Un ejemplo de repositorio institucional universitario: www.saber.ula.ve*. INCI v.31 n.1 Caracas Ven.
http://www.scielo.org.ve/scielo.php?pid=S0378-18442006000100007&script=sci_arttext
- DCMI. (2004). *Dublin core Metadata Element Set, Version 1.1.: Reference Description*. The Dublin Core Metadata Initiative
<http://dublincore.org/documents/usageguide/elements.shtml>
Recuperado el 10 de marzo de 2005
- (2006). *The Directory of Open Access Repositories - OpenDOAR*. University of Nottingham, UK and University of Lund, Sweden are developing the OpenDOAR service for Open Access to research information.
<http://www.opendoar.org/>
- Drake Miriam (2004). *A Institutional Repositories: Hidden Treasures*, Georgia Institute of Technology, *Searcher* [Revista electrónica] 12 (5)
<http://www.infotoday.com/searcher/may04/drake.shtml>
- Duke Monica (2003). "Delivering OAI Records as RSS: An IMesh Toolkit module for facilitating resource sharing". *Ariadne*[Revista electrónica] 37.
<http://www.ariadne.ac.uk/issue37/duke/>
- Ebbinghouse Carol (2005). "Open Access: The Battle for Universal, Free Knowledge" *Searcher*[Revista electrónica] 13 (3)
<http://www.infotoday.com/searcher/mar05/ebbinghouse.shtml>
- EPrints Handbook, *A guide to the EPrints repository software package, This manual is currently at version 0.91*.
<http://www.eprints.org/documentation/handbook/>
- Foster, N. and Gibbons, S. (2005). "Understanding Faculty to Improve Content Recruitment for Institutional Repositories", *D-Lib Magazine*, [Revista electrónica] 11 (1)

- <http://www.dlib.org/dlib/january05/foster/01foster.htm>
Gallagher Richard (2005). "Why We Need Institutional Repositories",
The Scientist [Revista electrónica] 10 (19)
<http://www.the-scientist.com/2005/10/10/8/1/>
- Gelfan Julia (2006). "7th International Conference on Grey Literature", *Library Hi Tech News* 23 (1) 13-16.
<http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&hdAction=Inkpdf&contentId=1541663>
- Gerstein, M. and J. Junker (2001). "Blurring the boundaries between scientific 'papers' and biological databases." *Nature* (7 May).
<http://www.nature.com/nature/debates/e-access/Articles/gerstein.html>
- Graham Rebecca A. (2001) "Metadata harvesting", *Library Hi Tech* , 19 (3), 290-295
<http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=861284>
(2005) *HaIRST, Institutional Repositories, metadata harvesting and global discovery services, a research approach to a development project*, Harvesting Institutional Resources in Scotland Tastbed
http://www.jisc.ac.uk/index.cfm?name=fairsynthesis_hairst
- Harnad, S. (2001). "The Self-Archiving Initiative." *Nature* 410. 1024-1025.
<http://www.ecs.soton.ac.uk/~harnad/tp/nature4.htm>
- Heery, Rachel, Anderson, Sheila (2005). *Digital Repositories Review UKOLN*, University of Bath Arts and Humanities Data Service.
http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf
- Henry, G. (2003). "On-line Publishing in the 21st Century: Challenges and Opportunities." *D-Lib Magazine* [Revista electrónica] 9(10)
<http://www.dlib.org/dlib/october03/henry/10henry.html>
- Hey, T. and A. E. Trefethen (2005). "Cyberinfrastructure for e-Science." *Science* 308(5723), 817-821.
<http://eprints.ecs.soton.ac.uk/10844/>
- Hoorn, E. (2005). "Repositories, Copyright and Creative Commons for Scholarly Communication." *Ariadne* [Revista electrónica] (45).
<http://www.ariadne.ac.uk/issue45/hoorn/>
- Houghton, J. W., C. Steele, et al. (2004).
"Research practices and scholarly communication in the digital environment."
Learned Publishing [Revista electrónica]17(3), 231-249.
<http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2004/00000017/00000003/art00009;jsessionid=r2fjvw0peq9p.victoria>
- Howto: Create a Data Repository*. By default EPrints is set up for archiving research output - articles, conference papers, technical reports, theses etc. However, it's straightforward to configure EPrints for different kinds of deposits, such as data or multimedia. Follow these steps to create a data repository
<http://www.eprints.org/software/howto/data/>
- Hunter, Philip (2005), *OAI and OAI-PMH for absolute beginners: a non-technical introduction [Introduction to OAI and Harvesting]* . Delivered at CERN workshop on Innovations in Scholarly Communication (OAI4), Geneva, Switzerland . Presentation.
<http://eprints.rclis.org/archive/00005512/01/OAI4-tutorial-hunter.pdf>
- Hurd, J. (2000). "The Transformation of Scientific Communication: A Model for 2020." *Journal of the American Society for Information Science* 51(14): 1279-1283.
<http://www3.interscience.wiley.com/cgi-bin/fulltext/74500450/PDFSTART>
- Jacob Inés, Joseba Abaitua, JosuKa Díaz, Fernando Quintana, Garikoitz Echebarria, (2004) *OAI-PMH. Estado del arte año 2004*. DELI.
http://www.deli.deusto.es/Resources/Documents/OAI-PMH_estado_del_arte.pdf
- JISC, Focus on Access to Institutional Resources (FAIR) Programme
http://www.jisc.ac.uk/index.cfm?name=programme_fair
- Information Environment JISC (2006). JISC Digital Repository Wiki.

- http://www.ukoln.ac.uk/repositories/digirep/index/JISC_Digital_Repository_Wiki
Johnson, Richard K. (2002). "Institutional repositories: partnering with faculty to enhance scholarly communication"; *D-Lib Magazine*, November 2002.
<http://www.dlib.org/dlib/november02/johnson/11johnson.html>
- Joint, Nicholas (2006). "Institutional Repositories, Self-Archiving and the Role of the Library." *Library Review* 55(2), 81-84.
<http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&hdAction=Inkhtml&contentId=1545768>
- Jones, Richard. (2004). "DSpace vs. ETD-db: Choosing software to manage electronic theses and dissertations", *Ariadne* [Revista electrónica] 38
<http://www.ariadne.ac.uk/issue38/jones/>
- Jones, R., Theo A. and MacColl, J. (2006), *The Institutional Repository*, Oxford. 250 pp. Review by Sally Rumsey at *Ariadne*, Issue 47 April 2006
<http://www.ariadne.ac.uk/issue47/rumsey-rvw/>
- Kling, Rob. (1999). "What is Social Informatics and Why Does it Matter?" *D-Lib Magazine* [Revista electrónica] 5(1).
<http://www.dlib.org/dlib/january99/kling/01kling.html>
- Kling, Rob. and McKim Geoffrey (1999). "Scholarly Communication and the Continuum of Electronic Publishing." *Journal of the American Society for Information Science* 50(10), 890-906.
<http://www.citebase.org/fulltext?format=application%2Fpdf&identifier=oai%3AarXiv.org%3Acs%2F9903015>
- Kling Rob, Howard Rosenbaum and Steve Sawyer, *Understanding and Communicating Social Informatics: A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*, Information Today, 2005, 216 pp.
- Lafuente, Antonio (2004). "Bien común y Open Access", *Madri+d* [Revista electrónica]
<http://www.madrimasd.org/InformacionIDI/Noticias/Noticia.asp?Buscador=OK&id=15498&Sec=2>
- Lagoze, Carl, and Herbert Van de Sompel (2001) "The Open Archives Initiative: Building a low-barrier interoperability framework." *Joint Conference on Digital Libraries*. Borrador
<http://www.cs.cornell.edu/lagoze/papers/oai-jcdl.pdf>
- Lagoze, Carl, and Herbert Van de Sompel (2003) "The making of the Open Archives Initiative Protocol for Metadata Harvesting", *Library Hi Tech*, Volume 21 Issue 2 2003 (pp. 118-128)
<http://www.emeraldinsight.com/Insight/viewContentItem.do?contentType=Article&contentId=861361>
- Lawrence, S. (2001). "Free online availability substantially increases a paper's impact" *Nature*, 31
<http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>
- López, Clara. (2005) *Los Repositorios de Objetos de Aprendizaje como soporte a un entorno e-learning*, Tesina doctoral, Universidad de Salamanca. (Director Francisco José García Peñalvo)
- Lynch, Clifford A., Joan K. Lippincott(2005). "Institutional Repository Deployment in the United States as of Early 2005", *D-Lib Magazine* [Revista electrónica] 11 (9)
<http://eprints.rclis.org/archive/00006368/>
- Lynch, Clifford A.(2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL Bimonthly Report* [Revista electrónica] 226.
<http://www.arl.org/newsltr/226/ir.html>
- MacKenzie Smith , et al, "Dspace. An Open Source Dynamic Digital Repository", *D-Lib Magazine*, January 2003. Volume 9 Number 1.
<http://dx.doi.org/10.1045/january2003-smith>
- McFarland Rob, Geri Short, *Creating an Institutional Repository Using DSpace*, CDT Washington University, St. Louis, Missouri, USA
http://library.wustl.edu/librarian/institutional_repository.ppt
- Medeiros, N. (2004). "Of budgets and boycotts: The battle over open access

- publishing." *OCLC Systems and Services* 20(1), 7-10.
http://eprints.rclis.org/archive/00001193/01/ELIS_OTDCF_v20no1.PDF
- Mellon Foundation (2006). *A Technology Analysis of Repositories and Services: A Proposal Submitted to the Mellon Foundation*, Johns Hopkins University, Sheridan Libraries.
<http://ldp.library.jhu.edu/projects/repository/documents/CNI-writeup.pdf>
<http://ldp.library.jhu.edu/projects/repository>
- OAI for Beginners- Overviews. O.-F. tutorial.
<http://www.oaforum.org/tutorial/index.php>
- OAI and OAI-PMH for absolute beginners: a non-technical introduction [Introduction to OAI and Harvesting]
<http://www.oaforum.org/tutorial/index.php>
- (2006). *Open Archives Initiative*, Wikipedia, the free encyclopedia
http://en.wikipedia.org/wiki/Open_Archives_Initiative
- Open Archives Initiative. *Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14*
<http://www.openarchives.org/OAI/2.0/guidelines.htm>
- The Open Archives Initiative Protocol for Metadata Harvesting Protocol*. Version 2.0 of 2002-06-14 Document Version 2004/10/12
<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- Odlyzko, A. (1995). "Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals." *International Journal of Human-Computer Interaction* 42(1), 71-122.
<http://www.dtc.umn.edu/~odlyzko/doc/tragic.loss.long.pdf>
- Odlyzko, A. (2002). "The rapid evolution of scholarly communication." *Learned Publishing* 15(1), 7-19.
<http://www.si.umich.edu/PEAK-2000/odlyzko.pdf&e=10401>
- Peek, Robin (2004). "Googling DSpace.(Focus on Publishing)(pilot project between Google and DSpace, digital repository of academic research)" *Information Today*; [Revista electrónica] 21 (6)
<http://www.infotoday.com/IT/jun04/index.shtml>
- Pinfield, Stephen (2005), "A mandate to self archive? The role of open access institutional repositories" *Serials* 18(1), March, 2005
http://eprints.nottingham.ac.uk/archive/00000152/01/mandate_to_archive.pdf
- Pinfield, Stephen, Mike Gardner, and John MacColl (2002) "Setting up an institutional e-print archive" *Ariadne* [Revista electrónica], 31.
<http://www.ariadne.ac.uk/issue31/eprint-archives/intro.html>
- Proberts, Steve y Jenkins, Celia (2006). "Documentation for Institutional Repositories", *Learned Publishing*, 19 (1).
<http://magpie.lboro.ac.uk/dspace/bitstream/2134/782/1/lppaper+-+final.pdf>
- Prom Christopher J. (2003). "Reengineering archival access through the OAI protocols", *Library Hi Tech*, 21 (2)
<http://wotan.openlib.org/doi/data/Articles/julfsernry:2003:v:21:i:2:p:199-209.html>
- Proyecto DIGIBUG de la Biblioteca Universitaria de Granada y otros proyectos de Repositorios Institucionales*
http://www.ugr.es/~biblio/digibug/e_ciencia.htm
- Rapple, Brendan (2003). "Scholarly journals in the digital age", *Contemporary Review Academic Research Library*; 283 (1652), 129.
<http://www.highbeam.com/doc/1G1:109569176/Scholarly+journals+in+the+digital+age.html?refid=SEO>
- Registro para políticas de acceso abierto de las instituciones repositorios institucionales
<http://www.uv.es/=barrueco/declaration.sp.html>
- ROARMAP (Registry of Open Access Repository Material Archiving Policies) as recommended by the Berlin Declaration
<http://www.eprints.org/openaccess/policysignup/>
- Registry of Open Access Repositories (ROAR)
<http://archives.eprints.org/>

- Shearer, Kathleen (2002). "The Open Archives Initiative: Developing an Interoperability Framework for Scholarly Publishing", *Canadian Association of Research Libraries*, Backgrounder Series No. 5.
http://www.carl-abrc.ca/projects/scholarly/open_archives.PDF
- Simeoni, Fabio (2004). "The Case for Metadata Harvesting", *Library Review*, 53 (5), 255 - 258
<http://hairst.cdlr.strath.ac.uk/documents/MetadataHarvesting.pdf>
- Suber, P. (2004). Open Access Overview.
<http://www.earlham.edu/~peters/fos/overview.htm>
- Sullivan Shirley , Lynne Horwood, Jane Garner, Eve Young, *Bringing hidden treasures to light: illuminating DSpace*, [presentación Microsoft PowerPoint], University of Melbourne
<http://www.vala.org.au/vala2004/2004pdfs/17SHGY.PDF><http://www.infodiv.unimelb.edu.au/ird/ird/presentations/valapresent2.pps>
- (2005). *Thomson scientific launches web citation index: new scholarly index directs researchers to valuable content from repositories* , *Citation Index: Institutional Repositories and Subject-Specific Archives*
<http://arrow.edu.au/docs/files/Thomson%20Scientific%2020051128.pdf>
- Unsworth, J. C., P. Courant, et al. (2005). *The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences* (for public comment), 64.
- Van der Kuil, Annemiek, Feijen, Martin (2004), "The Dawning of the Dutch Network of Digital Academic REpositories (DARE): A Shared Experience", *Ariadne* [Revista electrónica], 41
<http://www.ariadne.ac.uk/issue41/vanderkuil/>
- Van de Sompel, Herbert (2002). "Introductory Talk and Presentation of the OAI-PMH 2.0.", in *2nd Workshop on the Open Archives Initiative (OAI): Gaining Independence with E-Prints Archives and OAI*, CERN, 1, Geneva, Switzerland. Geneva[video].
<http://agenda.cern.ch/askArchive.php?base=agenda&categ=a02333&id=a02333s7t1/video>
- Ware, Mark (2004). "Institutional repositories and scholarly publishing", *Learned Publishing*, 17, (2), 115-124
<http://www.ingentaconnect.com/content/alpsp/lp/2004/00000017/00000002/art00006;jsessionid=ft9snsmqkhg.alice>
- Weitzman J. B. (2004) "(Mis)Leading Open Access Myths", Open Access Now. BioMed Central.
http://www.biomedcentral.com/openaccess/inquiry/myths/Open_Archives_Initiative

11. Anexos

ANEXO A. Requerimientos técnicos de Software

CDSware

Última versión:

CDSware v0.7.1 - 04 May 2005

Licencia:

GNU-GPL

Características principales:

** Colecciones con estructura arborescente para fácil navegación (documentos organizados en colecciones)

** Potente motor de búsqueda

- Índices especialmente diseñados para realizar búsquedas veloces en repositorios con más de millón y medio de registros.

- Interfaces de búsqueda simple y avanzada fáciles de adaptar a requerimientos particulares

- Combina búsquedas en metadatos, texto completo y citas en una sola consulta

- Resultados agrupados por colección.

** Metadatos flexibles

- Utiliza MARC 21

- Maneja artículos, libros, tesis, fotos, videos, objetos de museo y

más

- Ligas y despliegado de resultados modificables para requerimientos particulares

** Múltiples formatos de salida: HTML, XML, MARC y OAI

Requerimientos:

** Sistema Operativo tipo UNIX. Esta probado en Debian GNU/Linux, pero cualquier SO, que soporte el software listado a continuación, puede usarse para CDSware.

** 4.0.x < MySQL < 4.1.0 (cliente --localmente-- y servidor --puede estar remotamente--)

** Servidor web APACHE 2 > 2.0.43 con soporte para cargar módulos DSO

** Python >= 2.3.2 con los siguientes módulos:

- MySQLdb version 0.9.2

- Numeric module >= v21

- (recomendado) PyStemmer, para indexar y ranking

- (recomendado) PyRXP, para un procesamiento rápido de XML MARC

- (recomendado) Gnuplot.Py, para generar gráficos

- (opcional) 4suite, alternativa a PyRXP
- (opcional) Psycho, para optimizar la ejecución del código
- ** Modulo de APACHE mod_python 3.x
- ** PHP >= 4.3.0 compilado como módulo de APACHE incliyendo soporte para MySQL
- ** PHP compilado para ejecutar desde línea de comandos (CLI)
- ** WML - Website META Language >= 2.0.8
- ** pdftotext > 3.0 para poder extraer referencias y palabras de un documento a texto completo
- ** Para poder realizar búsquedas por palabras an archivos de texto completo:
 - antiword, catdoc, o wvText (para archivos de MS Word)
 - pdftotext o pstotext (para archivos PDF)
 - pstotext o ps2ascii (para archivos PostScript)
 - pptHtml y html2text (para archivos MS PowerPoint)
 - xlhtml y html2text (para archivos MS Excel)
- ** (opcional) RXP y 4suite si se eligió instalar el modulo de Python para XML MARC
- ** (recomendado) Gnuplot programa de trazado desde línea de comandos
- ** (recomendado) CLISP, SBCL o CMUCL para análisis del log del servdor web y el programa de chequeo de metadatos

Eprints

Última versión:

GNU EPrints v2.3.13.1

Licencia:

GNU-GPL

Principales características:

- Fácil configuración y adaptación a diversos requerimientos
- La version 2.3 puede considerarse lista para usarse, aunque si se desea adaptar mejor el software necesitara afinaciones en los archivos de configuración.

Requerimientos:

- ** SMBD MySQL (probado en la versión 3.23.29a-gamma)
- ** Servidor web APACHE (v1.3.14) con el modulo mod_perl (v1.25)
- ** Perl 5.6 y sus modulos:
 - Data::ShowTable, requerido por la interfaz de MySQL
 - DBI v1.14, requerido por la interfaz de MySQL
 - Msql-Mysql Module v1.2215
 - MIME::Base64 v2.11,el modulo Unicode::String lo requiere para su instalación
 - Unicode::String v2.06
 - XML::Parser v2.30

DSpace

Última versión:

DSpace Stable v1.3.2

Licencia:

BSD

Principales características:

- Puede utilizarse para almacenar, indexar, preservar y redistribuir material de investigación en formatos digitales.
- Es flexible puede ocuparse como software para repositorios institucionales, como manejador de registros electrónicos, repositorio de objetos de aprendizaje, y más
- Puede adaptarse y extenderse para satisfacer necesidades particulares

Requerimientos:

- Sistema Operativo tipo UNIX (Linux, HP/UX, etc.)
- Java Development Kit >= v1.4
- Apache Ant >= v1.5
- PostgreSQL >= v7.3 ó Oracle >= v9
- Jakarta Tomcat 4.x/5.x o un equivalente como Jetty o Caucho Resin

Fedora

Última versión:

Fedora v2.1b

Licencia:

ECL 1.0 (Educational Community License)

Principales características:

- Su flexibilidad le permite funcionar como repositorio digital para una variedad de casos de uso. Entre éstos están los gestores de activos digitales, repositorios institucionales, archivos digitales, sistemas gestores de contenidos, publicaciones escolares, bibliotecas digitales
- Fue creado teniendo en cuenta la complejidad del contenido digital: no solo existen documentos de texto. Puede manejar audio, video e imágenes
- Soporta repositorios distribuidos
- Integración fácil con otras aplicaciones y sistemas.

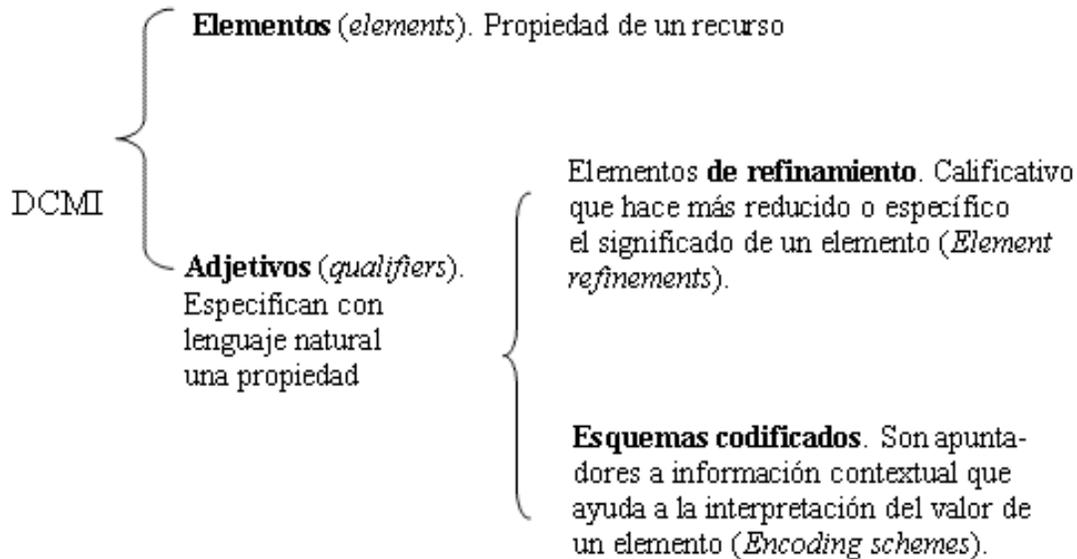
Requerimientos:

- Java Development Kit >= v1.4
- (optional) MySQL 4.x u Oracle 9i

ANEXO B. Dublin Core Qualified

I. Definiciones

La Iniciativa de Metadatos *Dublin Core Qualified* está formada por elementos (*elements*) y adjetivos (*qualifiers*). Los elementos son los metadatos de más alto nivel, definen las propiedades de un recurso. Los adjetivos especifican con lenguaje natural una propiedad, en ningún caso pueden utilizarse para extender su valor semántico. Los adjetivos pueden ser de dos tipos: refinamientos del elemento (*element refinements*) y esquemas codificados (*encoded schemes*). Los refinamientos son calificativos que hacen más reducido o específico el significado de un elemento. Los esquemas son apuntadores a información contextual que ayuda a la interpretación del valor de un elemento, están formados por reglas de análisis gramaticales, notaciones formales, estándares, etcétera.



II. Descripción de los elementos y sus adjetivos

title . Es el nombre dado a un recurso. Usualmente, *title* será el nombre bajo el cual el recurso es oficialmente conocido.

Refinamiento:

- **alternative** . Cualquier título usado como sustituto o alternativo al título oficial del recurso. Se pueden incluir abreviaciones del título o traducciones.

creator. Es la entidad responsable de la creación del contenido del recurso, puede ser una persona (autor), organización o un servicio.

Adjetivos: no tiene.

subject. Se refiere al tema del contenido de un recurso.

Esquemas:

- **LCSH**. Library of congress Subject Headings.
- **MESH**. Medical Subject Headings. <http://www.nlm.nih.gov/mesh/meshhome.html>.
- **DDC**. Dewey Decimal Classification. <http://www.oclc.org/dewey/index.htm>.

- **LCC.** Library of Congress Classification. <http://lcweb.loc.gov/catdir/cpsol/lcco/lcco.html>.
 - **UDC.** Universal Decimal Classification. <http://www.udcc.org>.
- description.** Describe del contenido del recurso.
Refinamientos:
- **tableOfContents** . Una lista de subunidades del contenido del recurso.
 - **abstract** . Resumen del contenido del recurso.
- publisher** . Es la entidad responsable de que el recurso esté disponible. Puede ser una persona, una organización o un servicio.
Adjetivos: no tiene.
- contributor.** Es una entidad responsable de contriciones en el contenido del recurso. Puede ser una persona, una organización o un servicio.
Adjetivos: no tiene.
- date** . Una fecha asociada con un evento en el ciclo de vida del recurso.
Refinamientos:
- **created** . Fecha de creación del recurso.
 - **valid** . Fecha de validez del recurso (usualmente un periodo).
 - **available** . Fecha en que el recurso se hará o se hizo disponible.
 - **issued** . Fecha formal de publicación.
 - **modified** . Fecha en la cual el recurso fue modificado.
 - **dateAccepted** . Fecha de aceptación del recurso (tesis, artículos, etc.).
 - **dateCopyrighted** . Fecha de declaración de copyright.
 - **dateSubmitted** . Fecha de solicitud de revisión (tesis, artículo).
- Esquemas:
- **period** . Especificación de los límites de un intervalo de tiempo. <http://dublincore.org/documents/dcmi-period/>.
 - **W3CDTF** . Reglas de codificación de la w3c para fechas y horas, basado en ISO 8601. <http://www.w3.org/TR/NOTE-datetime>.
- type** . La naturaleza o el género del recurso
Esquemas (vocabulario):
- **collection** . Una colección es una suma de objetos. El recurso es descrito como grupo, sus partes pueden ser descritas y navegables de manera individual.
 - **dataset** . Es información codificada es una estructura definida (listas, tablas y bases de datos), pretendiendo ser útiles para procesamiento directo de una máquina.
 - **event** . Es una ocurrencia no persistente, temporal. Los metadatos proveen información para identificar el objetivo, lugar, duración, responsables y enlaces relacionados con el evento. El recurso tipo *event* podría no ser recuperable si su instanciación ha expirado o no ha ocurrido. (p.e. exhibiciones, conferencias, *workshop* , presentaciones, batallas, bodas, etcétera).
 - **image** . Una imagen es una representación visual que no sea texto. (fotografías, pinturas, dibujos gráficos, animaciones, películas).
 - **interactiveResource** . Es un recurso que requiere interacción del usuario para entenderse, ejecutarse o sentirse. (páginas web, multimedia, chats, realidad virtual).
 - **service** . Sistema que provee uno o más funciones con valor para el usuario final (servicio de fotocopia, servicio bancario, servicio de autenticación, préstamo interbibliotecario, z39.50 o web server).
 - **software** . Programa de computo en fuente o compilado, que puede estar disponible para instalación.
 - **sound** . Recurso que principalmente se reconoce como audio (música, discurso, sonido).
 - **text** . Recurso cuyo contenido es principalmente palabras para leer (libros, cartas, disertaciones, poemas, periódicos, artículos). Facsímiles e imágenes de textos son de tipo *text* .
 - **physicalObject** . Objeto inanimado tridimensional (computadora, pirámide, escultura). Representaciones o sustitutos digitales deben usar *image*, *texto* o alguno de los otros *type* .

format . La manifestación física o digital del recurso. Usualmente, puede incluir el medio o las dimensiones del recurso. Puede ser usado para determinar el software, hardware u otro equipo necesario para desplegar u operar el recurso.

Refinamientos:

- **extent** . El tamaño o la duración del recurso.
 - **medium** . El material o portador físico del recurso.
- Esquemas
- **IMT** . Internet Media Type. www.isi.edu/in-notes/iana/assignments/media-types/media-types.
 - **MIME** . <http://dublincore.org/usage/terms/references/#MIME>.

identifier . Una referencia no ambigua hacia el recurso dentro de un contexto dado. Se recomienda utilizar una cadena o número autogenerado.

Esquemas:

- **URI** . Uniform Resource Identifier. www.ietf.org/rfc/rfc2396.txt.

source . Referencia a un recurso del cual se derivó el recurso que se describe.

Esquemas:

- **URI** . Uniform Resource Identifier. www.ietf.org/rfc/rfc2396.txt.

language. Idioma del contenido intelectual del recurso.

Esquemas:

- **ISO639-2** . Códigos de representación de nombres de idiomas. www.loc.gov/standards/iso639-2/langhome.html.
- **RFC1766** . Internet RFC 1766 etiquetas para la identificación de idiomas, especifica un código de 2 letras tomado del iso 639, seguido opcionalmente de 2 letras del país tomadas del código iso 3166. www.ietf.org/rfc/rfc1766.txt.
- **RFC3066** . Internet RFC 3066 etiquetas para la identificación de idiomas. Especifica código de dos letras para la parte 1 o un código tres letras para la parte 2, tomados del ISO 639. Este RFC reemplaza al RFC1766.

relation . Referencia un recurso relacionado con el recurso descrito.

Refinamientos:

- **isVersionOf** . El recurso descrito es versión, edición o adaptación del recurso referenciado. Cambios en la versión implica cambios sustantivos en contenido pero no en formato.
- **hasVersion** . El recurso descrito tiene al recurso referenciado como versión, edición o adaptación.
- **isReplacedBy** . El recurso descrito es sustituido, desplazado o reemplazado por el recurso referenciado.
- **Replaces** . El recurso descrito suplanta, reemplaza el recurso referenciado.
- **isRequiredBy** . El recurso descrito es requerido por el recurso referenciado, física o lógicamente.
- **Requires** . El recurso descrito requiere el recurso referenciado para apoyar o sustentar su función, entrega o coherencia.
- **isPartOf** . El recurso descrito es física o lógicamente parte del recurso referenciado.
- **hasPart** . El recurso descrito incluye el recurso referenciado física o lógicamente.
- **isReferencedBy** . El recurso descrito es referenciado, citado o direccionado por el recurso referenciado.
- **references** . El recurso descrito referencia, cita o apunta al recurso referenciado.
- **isFormatOf** . El recurso descrito tiene el mismo contenido intelectual del recurso referenciado, pero presentado en otro formato.
- **hasFormat** . El recurso descrito preexiste al recurso referenciado, es el mismo contenido intelectual presentado en otro formato.
- **conformsTo** . Referencia a un estándar establecido al que pertenece el recurso.

Esquemas:

- **URI** . Uniform Resource Identifier. www.ietf.org/rfc/rfc2396.txt.

coverage . La extensión o alcance del contenido del recurso. Regularmente se incluye una ubicación de espacio, lugar, coordenadas geográficas; también puede ser un periodo de tiempo, una fecha, un rango de fechas.

Refinamientos:

- **spatial** . Características de espacio geográfico del contenido intelectual del recurso.

Esquemas:

- **Point** . Identifica un punto en el espacio utilizando sus coordenadas geográficas. c.
 - **ISO3166** . Código para representar los nombres de países.
<http://www.din.de/gremien/nas/nabd/iso3166ma/codlstp1/index.html>
 - **Box** . Identifica una región del espacio utilizando sus límites geográficos.
<http://dublincore.org/documents/dcmi-box>.
 - **TGN** . The Getty Thesaurus of Geographic Names.
<http://www.getty.edu/research/tools/vocabulary/tgn/index.html>.
- **temporal** . Características de tiempo, época, del contenido intelectual del recurso.
- Esquemas:
- **Period** . Especificación de los límites de un intervalo de tiempo.
<http://dublincore.org/documents/dcmi-period/>.
 - **W3CDTF** . Reglas de codificación de la w3c para fechas y horas, basado en ISO 8601. <http://www.w3.org/TR/NOTE-datetime>.

rights. Información sobre la propiedad de recurso y su manejo en materia de los derechos de autor. Puede incluir una leyenda, o una referencia hacia otro sitio que provea las políticas de uso.

Adjetivos: no tiene.

audience . El tipo de sector o grupo hacia quienes el recurso está orientado, esto lo puede determinar el autor, el editor o una tercera parte.

Refinamientos:

- **mediator** . Una entidad que consigue el acceso al recurso y para la cual está orientado o es útil el recurso. Ultimo beneficiario del recurso o una entidad que consigue el acceso.
- **educationLevel** . Audiencia en términos de su progreso en un contexto educativo.

III: Referencias

DCMI. (2004). Dublin Core Metadata Element Set, Versión 1.1: Reference Description.

Disponible en <http://dublincore.org/documents/dces/>.

DCMI (2005). DCMI Metadata Terms. Disponible en <http://dublincore.org/documents/dcmi-terms/>.

DCMI (2004). DCMI Type Vocabulary. Disponible en <http://dublincore.org/documents/dcmi-type-vocabulary/>.