

Acceso Abierto al conocimiento científico, repositorios digitales y adopción de estándares desde el SIU.

Emiliano Marmonti^a, Isabel B. Piñeiro^a, María de Luján Gurmendi^a

^a Consorcio SIU, Santa Fé 1548 P.11 Frente. Ciudad Autónoma de Buenos Aires. República Argentina.

emarmonti@siu.edu.ar , ipineiro@siu.edu.ar , lujan@siu.edu.ar

Resumen. En esta ponencia se tratarán los aspectos conceptuales básicos y antecedentes del acceso abierto al conocimiento científico, el surgimiento de Repositorios Institucionales, la situación de los mismos en América Latina, particularmente en Argentina, el surgimiento de las redes de Repositorios a nivel nacional e internacional, los aspectos clave de interoperabilidad para la generación y funcionamiento de dichas redes, aspectos tecnológicos relacionados con metadatos, Open Data y la web semántica, y el papel que el SIU ha desempeñado en la República Argentina en cooperación con instituciones de orden nacional en la faceta de referente técnico, facilitador y proveedor de soluciones tecnológicas en sintonía con el paradigma de su creación y funcionamiento, relacionado con la provisión de información a partir del énfasis en la calidad de los datos.

Palabras Clave: Acceso Abierto, Open Access, Repositorios Institucionales, Interoperabilidad, Metadatos, Web Semántica, Infraestructuras para Ciencia Abierta, Harvesting, Cosecha de datos.

1. Antecedentes

El SIU – Sistema de Información Universitario

El SIU (<http://www.siu.edu.ar>) es un consorcio de 45 Universidades Nacionales Públicas creado en el ámbito de la Secretaría de Políticas Universitarias del Ministerio de Educación de la Nación.

El objetivo del SIU es dotar al sistema universitario nacional público de herramientas tecnológicas que colaboren con la gestión, brinden servicios a los ciudadanos universitarios y simultáneamente produzcan datos de calidad que sirvan para la toma de decisiones y análisis institucional.

Estos objetivos se logran en un ámbito de colaboración y participación activa de todos los actores involucrados. Actualmente existe una comunidad de unas 10.000 personas que participan en distintas Comunidades de Práctica que se corresponden con distintas temáticas.

En particular, el SIU Bibliotecas (<http://www.siu.edu.ar/siu-bibliotecas>) ha generado las condiciones para compartir la información existente en las bibliotecas universitarias y de otras instituciones que se han sumado, con el objetivo de propiciar el trabajo colaborativo y facilitar a los lectores el acceso a la información.

En 1998 comenzó la creación de la Base de Datos Unificada – BDU (<http://bdu.siu.edu.ar>), con la finalidad de reunir registros bibliográficos de las distintas bibliotecas universitarias. Esta iniciativa permite, que los registros sean “mapeables” desde los distintos formatos utilizados en las bibliotecas y a su vez se generó el software que permite que las bibliotecas importen esos registros a sus propias bases de datos, evitando la duplicación del trabajo de catalogación.

La BDU reúne 2.260.000 registros de 80 bibliotecas de diferentes instituciones (abril 2012). Este proceso ha sido acompañado con capacitación permanente para los bibliotecarios, en actualización sobre el uso de las TICs, incentivando el trabajo colaborativo entre las bibliotecas universitarias.

En el año 2004 el SIU realizó el primer taller sobre Bibliotecas Digitales, en coordinación con UNESCO e IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia). A partir del año 2009, el SIU comienza a trabajar, utilizando diversas herramientas en lo que denominamos la BDU2 (<http://bdu.siu.edu.ar/bdu2>).

El objetivo de la BDU2 es reunir recursos de información de valor académico de libre disponibilidad para el usuario final. Dichos contenidos pueden ser Tesis de diversos grados académicos, Artículos de Publicaciones Periódicas, Libros electrónicos, Material de alto valor Histórico digitalizado, Legislación Educativa, Videos, Entrevistas y mucho otro material puesto a disposición por instituciones académicas nacionales.

2. Conceptos relacionados con el proyecto

Acceso Abierto.

Open Access o Acceso Abierto, según Suber [1], “es aquél que permite el libre acceso a los recursos digitales derivados de la producción científica o académica sin barreras económicas o restricciones derivadas de los derechos de copyright sobre los mismos. Esta producción engloba no sólo artículos publicados en revistas, sino también otro tipo de documentos como objetos de aprendizaje, imágenes, datos, documentos audiovisuales, etc. El acceso es online a través de Internet, y salvo limitaciones tecnológicas y de conexión a la red del usuario, no debería estar restringido por otro tipo de imposiciones”.

Surge como Iniciativa, a partir de las Declaraciones de Budapest [2] y de Bethesda [3], aunadas en la Declaración de Berlín [4] que ha conseguido importantes adhesiones, entre otras de 373 instituciones entre las que se encuentran universidades de diversos países. Además UNESCO y Unión Europea tienen políticas para promover el Acceso Abierto, con recomendaciones sobre “buenas prácticas”.

La Declaración de Budapest sostiene que “La literatura que debería accesarse libremente en línea, es aquella que los académicos dan al mundo sin la expectativa de recibir pago...Por "acceso abierto" a esta literatura queremos decir su disponibilidad gratuita en Internet público, permitiendo a cualquier usuario leer, descargar, copiar, distribuir, imprimir, buscar o usarlos con cualquier propósito legal, sin ninguna barrera financiera, legal o técnica, fuera de las que son inseparables de las que implica acceder a Internet mismo. La única limitación en cuanto a reproducción y

distribución y el único rol del copyright en este dominio, deberá ser dar a los autores el control sobre la integridad de sus trabajos y el derecho de ser adecuadamente reconocidos y citados”.

Jānis Kārklīņš, en su prólogo [5] sostiene que “Como se ha indicado en su Constitución, la UNESCO se dedica a "mantener, incrementar y difundir el conocimiento". Por lo tanto, parte de su misión es construir sociedades del conocimiento, fomentando el acceso universal a la información y el conocimiento a través de tecnologías de la información y la comunicación” y en ese marco dice “Creo que este documento global será útil en términos de toma de decisiones y formuladores de políticas en los planos nacional e internacional” se destaca la finalidad de los repositorios digitales de maximizar la visibilidad e impacto, y optimizar el seguimiento y la gestión de la investigación, además de atender el imperativo de preservación de la información. Por otra parte también cuando se alojan materiales educativos es un apoyo para el aprendizaje.

Repositorios Institucionales

Lima Leite [6] indica que un repositorio institucional de acceso abierto es un servicio de información científica en un ambiente digital e interoperable dedicado a albergar organizadamente la colección formada por la producción científica o académica de una institución. Contempla, por lo tanto, la reunión, almacenamiento, descripción, organización, preservación, instrumentos de búsqueda y recuperación y sobre todo, la ampliación de la difusión de la información científica producida en la institución.

Esta última circunscripción del dominio de la producción a lo institucional diferencia los Repositorios Institucionales de Repositorios Digitales, Repositorios Temáticos, Repositorios de Datos, etc.

Metadatos e Interoperabilidad en Repositorios Institucionales.

Gómez Dueñas [7] define la interoperabilidad en Sistemas de Información documental como “La capacidad de un sistema de información para comunicarse y compartir datos, información, documentos y objetos digitales de forma efectiva (con una mínima o nula pérdida de su valor y funcionalidad), con uno o varios sistemas de información (siendo generalmente estos sistemas completamente heterogéneos, distribuidos y geográficamente distantes), mediante una interconexión libre, automática y transparente, sin dejar de utilizar en ningún momento la interfaz del sistema propio”.

Acompañando los contenidos digitales que se depositan en los repositorios, se incluye información descriptiva de los mismos, denominada comúnmente metadatos. García Martínez [8] define metadatos como “datos asociados a documentos que ayudan a los usuarios potenciales a tener un conocimiento anticipado de la existencia característica de los mismos; todo ello teniendo en cuenta que soportan una gran variedad de operaciones y que un usuario puede ser una persona o un programa”

Existen múltiples convenciones o esquemas de metadatos: Hallamos estándares para disciplinas tan variadas como el agro, la salud, recursos naturales como

acuíferos, biología, defensa y muchas otras. Sin embargo, casi la totalidad de los repositorios institucionales admite exponer sus metadatos en un esquema conocido como Dublin Core. El mismo posee un conjunto reducido de campos (entre los que figuran los autores, el título, la fecha de publicación, lugar, país, derechos por sobre el o los objetos digitales, entre otros). Si bien, lo anterior representa interoperabilidad sintáctica, la interoperabilidad semántica se daría desde la misión de normalización de los contenidos de estos metadatos a través de la intervención de un profesional de la información en la aplicación de las normas RCA2R o RAD para el área descriptiva de los mismos o la aplicación de esquemas conceptuales como ontologías, de herramientas o de herramientas que permitan una normalización terminológica tales como taxonomías y tesauros.

La interoperabilidad semántica es uno de los grandes desafíos a enfrentar en el caso de los repositorios institucionales.

Una gran parte de los Repositorios Institucionales contemplan un esquema de interoperabilidad¹ estructural basado en el protocolo OAI-PMH. El mismo es un protocolo basado en el protocolo http y lenguaje de marcado XML, el cual permite a través de un conjunto simple de comandos o verbos (pueden verse más detalles de los mismos en la Figura 1), exponer los metadatos de los objetos digitales (dentro de los cuales está la URL del recurso) y así facilitar un proceso denominado “cosecha” o en inglés, “harvesting” el cual implementan muchas de las redes para la construcción de portales de búsqueda producción intelectual de acceso abierto, un ejemplo de las mismas podemos verlo en la Figura 2, el portal de la Comunidad Económica Europea que cosecha datos de 328 repositorios institucionales.

¹ Cabe destacar que es parte de la definición de Repositorio Institucional, que el mismo sea interoperable. En el texto se destaca que comúnmente este aspecto funcional es instrumentado a través del protocolo OAI-PMH.

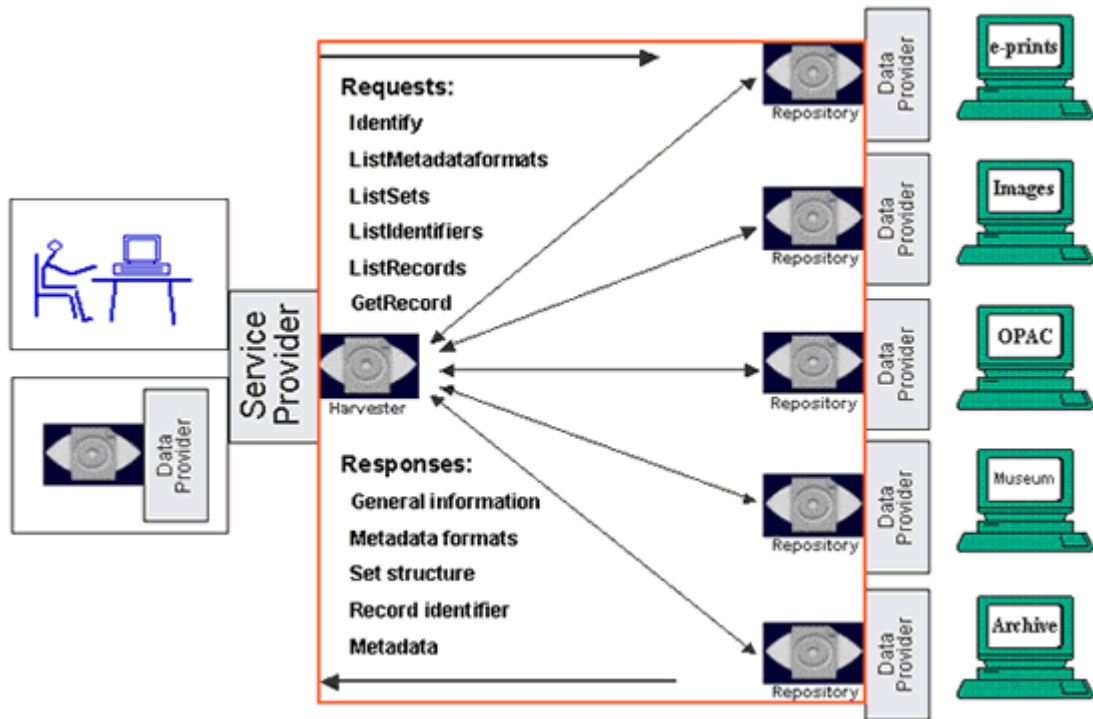


Fig 1 Ejemplo de los verbos OAI-PMH y la diferenciación de roles entre los Data Providers (quienes exponen sus metadatos) y Service Providers (quienes los cosechan y proveen un servicio de búsqueda para los mismos).

Digital Repository Infrastructure Vision for European Research
 MyDriver
 email password
 BUSQUE EN LOS REPOSITORIOS SABER MÁS DE DRIVER BUSCAR SOPORTE REGISTRE SU REPOSITORIO
 Inicio | Buscar | Navegar | Colecciones Repositorios
 Buscar en más de 5.810.000 documentos en 328 repositorios de 43 países
 Búsqueda Avanzada...
 No hay Anuncios.
 European Countries
 Specialized Communities
 Subject classification
 Ver todos
 Cree sus colecciones...
¿Es usted un administrador de repositorio?
 ¿Está su repositorio registrado en DRIVER? Si no, visite nuestro sitio dedicado "Registre su Repositorio", donde podrá registrar su repositorio con unos pocos clicks.
¿Podemos ayudar a su organización?
 El software de Driver (D-NET v 1.0) puede ser usado para poner en marcha portales similares para recolectar y agregar contenido desde varias fuentes. Provee búsqueda, navegación y servicios personalizados usando interfaces de usuario.
 Examine dos de las aplicaciones de ejemplo, el portal Belgium DRIVER y el portal de publicaciones españolas Recolecta .
 ¿Tiene preguntas o necesita más información?. Por favor, póngase en contacto con nuestro helpdesk y estaremos encantados de responderle.
 node3

Fig. 2 Ejemplo de una de las redes de repositorios y proveedor de servicio OAI-PMH, el portal DRIVER de la Unión Europea.

3. Contexto y ámbito del proyecto.

MINCYT – SNRD

En Argentina, el Ministerio de Ciencia, Tecnología e Innovación Tecnológica, impulsa la legislación para instituir la obligatoriedad del depósito de la producción científica financiada con fondos públicos. Ha creado el Sistema Nacional de Repositorios Digitales – SNRD (<http://repositorios.mincyt.gob.ar>) para estimular la creación de los mismos y establecer recomendaciones para hacerlos interoperables y accesibles.

El Proyecto de Ley presentado al Congreso se titula Creación de Repositorios Digitales Abiertos de Ciencia y Tecnología. CIENCIA ABIERTA ARGENTINA 2010 y propone en su Art. 1º. "...Los organismos e instituciones públicas que componen el Sistema Nacional de Ciencia, Tecnología e Innovación (SNCTI) y que reciben financiamiento del Estado Nacional, deberán desarrollar repositorios digitales institucionales de acceso abierto, propios o compartidos, en los que se depositará la

producción científico tecnológica resultante del trabajo, formación y/o proyectos, financiados con fondos públicos, de sus investigadores, tecnólogos, docentes, becarios de postdoctorado y estudiantes de maestría y doctorado...”

A través del Sistema Nacional de Repositorios Digitales - SNRD del MINCYT se han estudiado y consensado los estándares que hacen interoperables los Repositorios Institucionales y cuyo cumplimiento es imprescindible para adherir a dicho sistema. El SIU participa en el Comité de Expertos de Repositorios Digitales en general y en particular en las comisiones de metadatos e interoperabilidad.

Situación en América Latina de los Repositorios Institucionales.

Como se puede apreciar en la figura 3, la cantidad de repositorios institucionales en América Latina comparativamente con América del Norte o Europa, aún es pobre, pero ha incrementado fuertemente su número (tanto como volumen y contenido) en los últimos años.

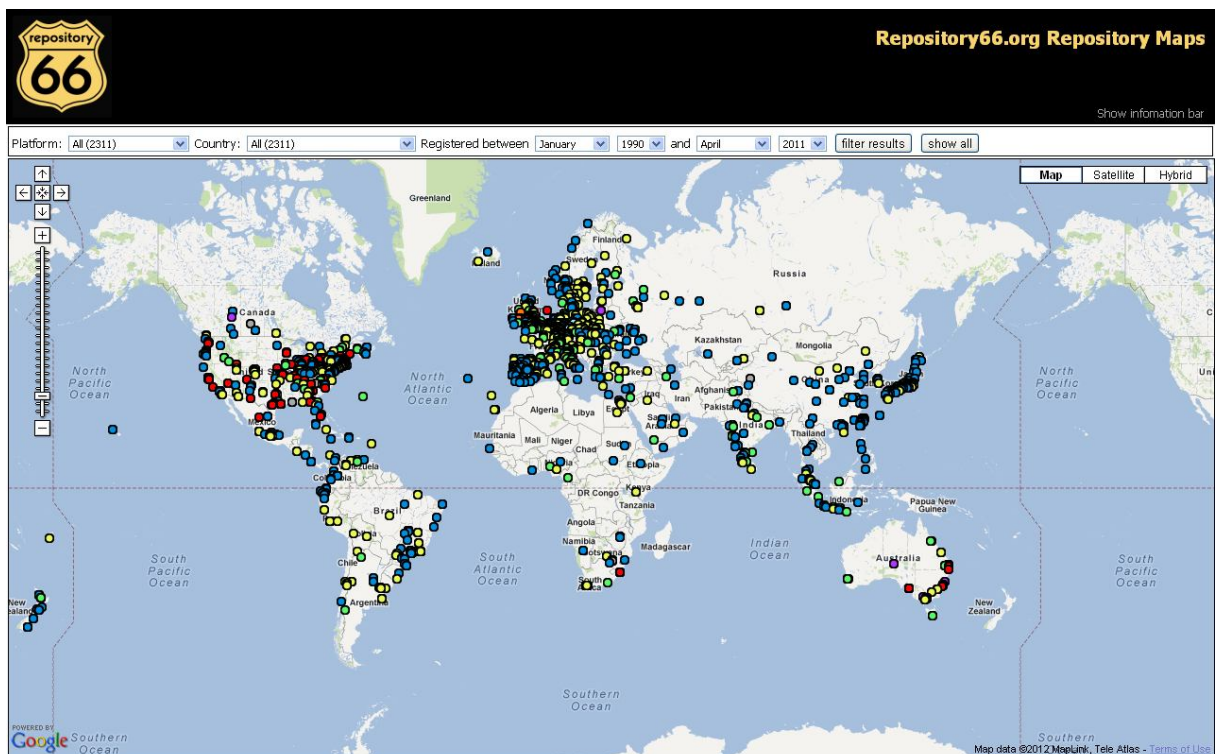


Fig. 3 Distribución de repositorios institucionales en el Mundo. Fuente <http://repository66.org>

Es importante destacar que en los últimos años han proliferado asimismo las redes de repositorios. Podríamos destacar entre ellas a redes generadas a partir de áreas geográficas y/o nivel académico, tales como redes de repositorios latinoamericanos (LaReferencia, patrocinada por la Red CLARA y el BID), en el área de ciencias sociales CLACSO-REDALYC, Recolecta en el caso de las Universidades Españolas, el portal DRIVER para la información científica de la Unión Europea, o redes de orden temático, tales como CIARD RING en el ámbito agropecuario, las pre-existentes redes en salud tales como la red Scielo de artículos científicos en áreas de salud de acceso abierto, u OCEANDOCs, en el caso de información de investigaciones relacionadas con los océanos (generado a partir de los aportes de instituciones internacionales). Por lo general, las redes de orden temático incluyen subdivisiones de orden geográfico y viceversa.

Las redes de repositorios, sean temáticas o generadas por una división geográfica aprovechan una de las características funcionales de los repositorios institucionales: La interoperabilidad.

Situación de los Repositorios Institucionales en Argentina.

Desde las instituciones que financian la investigación se van generalizando las políticas que establecen la obligatoriedad de depositar las publicaciones científicas (producidas con fondos públicos o provenientes de esas instituciones) en repositorios institucionales o temáticos de Acceso Abierto, reconocidas como “buenas prácticas” en las recomendaciones de la UNESCO.

En la actualidad se pueden hallar alrededor de 18² repositorios institucionales activos en la Argentina. Algunos de ellos contando con “mandato institucional”, es decir la obligatoriedad de realizar depósito de la producción intelectual por parte de los integrantes de esa institución, otros realizando tareas de marketing y sensibilización por parte de los productores de contenidos acerca de los beneficios tangibles que los mismos ofrecen.

En muchos de los casos, tanto de Latinoamérica como en los casos del exterior, incluso de las organizaciones internacionales, los repositorios son administrados por las Bibliotecas o por división equivalente en la institución. Adicionalmente, algunos repositorios implementan un proceso conocido como autoarchivo, el cual socializa el proceso de ingesta de los objetos digitales y sus metadatos depositando este proceso al menos inicialmente, en manos de los mismos productores de contenido. Por estas razones, las Bibliotecas se enfrentan a nuevos desafíos de orden tecnológico, de necesidades de capacitación y a un escenario donde la generación de la información no es centralizada (como lo era en los tradicionales procesos de catalogación) y pasa a ser su responsabilidad la normalización de la misma para posibilitar la recuperación de los contenidos y habilitar la interoperabilidad desde otras dimensiones adicionales a la sintáctica o estructural.

² Se indica una cantidad aproximada de repositorios, dado que hay iniciativas en fase de lanzamiento o iniciando su etapa de producción, aún no listadas en los directorios internacionales que registran repositorios.

En la investigación llevada a cabo por Miguel, Gómez y Bongiovani se habla en números del estado del “Acceso abierto real y potencial de la producción científica argentina. Los resultados de este estudio muestran que, en promedio, y para el conjunto de las temáticas estudiadas, el 70% de la producción científica argentina del período 2008-2010 incluida en *Scopus* se publica en revistas que se adhieren de una u otra forma al movimiento OA real y potencial”³. A través de este estudio se concluye que en Argentina existen condiciones favorables para el crecimiento de los repositorios con literatura científica generada bajo la modalidad de acceso abierto. [9]

4. Rol del SIU en la realidad de los Repositorios Institucionales en Argentina.

El SIU adhiere a la corriente de Acceso Abierto y ha impulsado la creación de Repositorios Institucionales en las universidades nacionales con acciones como el Taller sobre Bibliotecas Digitales realizado en el año 2004. A partir de esa fecha se advierte el surgimiento de repositorios institucionales que favorecen la visibilidad de la producción académica de las universidades nacionales e instituciones de investigación.

En el año 2009 se observa la puesta en producción de 7 repositorios institucionales académicos activos: (Biblioteca Digital de la Universidad Nacional de Cuyo, Memoria académica – FaHCE UNLP, SeDiCI-UNLP, CLACSO, Biblioteca Digital de la UNS, Repositorio Hipermedial de la UNR, Cartapacio de Derecho-UNICEN). Por otra parte, desde la puesta en producción de la BDU, la cual sólo contenía información referencial, es que los usuarios (principalmente los usuarios no-bibliotecarios) pretendían acceder al texto completo de las bases de datos bibliográficas, el concepto mismo de una base de datos referencial con el advenimiento de iniciativas como Google Books, fue perdiendo valor.

A medida que avanzan en el crecimiento los repositorios digitales, las instituciones comienzan a comprender la necesidad de respetar los protocolos de intercambio que permiten reconocer y “cosechar” la información de forma tal de recuperar los objetos digitales alojados y comienzan a implementar el protocolo OAI-PMH (Open Archive Initiative – Protocol Metadata Harvesting). El SIU comenzó entonces a trabajar, utilizando diversas herramientas open source sobre dos componentes principales: el cosechador OAI y el portal de acceso público a los recursos, para que a través de una única búsqueda se localice el objeto digital en cualquiera de los repositorios. Por otra parte se trabajó individualmente con las proveniencias, mostrando la necesidad de acordar los estándares para los metadatos de forma de hacer interoperables todos los repositorios institucionales y colaborando para la elaboración de la propuesta de directrices a partir de la experiencia de haber abordado todas las alternativas que se habían utilizado en la descripción de los objetos digitales.

En las figuras 4 y 5, se pueden apreciar algunas de las pantallas de la aplicación BDU2, la cual se halla en producción desde el año 2009.

³ Las temáticas analizadas en la investigación son medicina, física y astronomía, agricultura y ciencias biológicas, y ciencias sociales y humanidades.

Ayuda

Ingrese la expresión de búsqueda en la caja de texto señalada con la flecha azul.

Los resultados que la búsqueda arroje serán categorizados por diferentes criterios.

Si desea buscar un término truncado puede hacerlo agregando a su expresión el símbolo \$
Por ejemplo: democrac\$
Con lo que aparecerán los términos:
Democracia
Democráticos
Democráticas, etc.

Búsqueda

Expresión a Buscar:

Criterio:



Fig 4 Pantalla inicial de BDU2, nube de descriptores de temáticas.
<http://bdu.siu.edu.ar/cgi-bin/query.pl>

Tipo Material

Artículo (10)
Tesis (6)
Publicación periódica (2)
Tesis de Postgrado (2)
legislation (2)
Working Paper (1)

Autores

Castagnino, Juan M. (3)
Baque, Laura Cecilia (1)
Barbosa, Leonardo M. Q. (1)
Borchez, Carlos (1)
Bregni, Carlos (1)
Bugallo, Alicia I. (1)
Buteler, Micaela (1)
Cabrera, Ricardo (1)
Carlucci, Adriana M. (1)
Casabella, Adriana N. (1)
Castagnino, Juan Miguel (1)
Cossio Torres, Patricia (1)
Curiale, Carlos, J. (1)
Díaz-Barriga Martínez, Fernando (1)
Dominguez Cortinas, Gabriela (1)

Materias

NANOTECNOLOGIA (9)
FISICA (3)

Búsqueda

Expresión a Buscar:

Criterio:

ENVIAR CONSULTA

Cantidad de Resultados: 23
Páginas: 1 2

Ordenar por:

Nanotecnología Técnicas nanomagnéticas

Autor/es (contribuyente): Castagnino, Juan Miguel
Editor/Título Revista: Acta bioquímica clínica latinoamericana
Tipo de Material: journal article

Haga clic: aquí para acceder al Objeto Digital.

Repositorio: RD-Scielo Argentina ★★

Compartir en redes sociales

Herramientas ▾

Nanotecnología, microchips y microarreglos

Fig 5 Presentación de los resultados de una búsqueda sobre la información cosechada de los repositorios. Discriminación por Autores, por materias, por año, por Repositorio de proveniencia.

Web semántica y los repositorios institucionales.

Entre las misiones y funciones de SIU Bibliotecas, se incluye la colaboración con otras áreas del SIU para la aplicación de herramientas de control y unificación de terminologías.

Se han realizado eventos tales como un curso de tecnologías para la web semántica dictado por la Doctora Eva Méndez en el año 2009, se ha participado en un proyecto de investigación aplicando ontologías a la expresión de competencias educacionales en la conformación de una carrera universitaria, y en el año 2011, se realizó un experimento introduciendo tecnologías tales como Linked Open Data en la interfaz de consulta de BDU2. Dicho experimento fue demostrado en oportunidad de la Jornada Virtual de Acceso Abierto organizada por el Ministerio de Ciencia y Tecnología e Innovación Productiva y la Organización Panamericana de la Salud (OPS) [10].

Dicha interfaz experimental propone a partir de los resultados de una consulta, un conjunto de términos extraídos de la DBPedia⁴, los cuales se considera que pueden tener una relación semántica humanamente delimitada a través de la ontología que rige dicha herramienta.

Así en la Figura 6 puede apreciarse que ante una consulta del término Antibiótico, se ofertan los resultados correspondientes a los repositorios institucionales y adicionalmente se conecta dicho término de búsqueda con DBPedia, obteniéndose términos que han sido relacionados por los usuarios que han volcado en forma cooperativa dicho contenido. Así es como a partir de un término de búsqueda se puede conectar con conceptos semánticamente relacionados con el volcado originalmente.

The screenshot displays the BDU2 experimental interface. On the left, there are filter buttons for 'Autores', 'Materias', 'Fecha', and 'Repositorios', along with a green button 'Podría interesarle...'. Below this is a list of related terms with 'W' and 'bdu²' icons, including Azitromicina, Gatifloxacina, Flucloxacilina, Norfloxacino, Bencilpenicilina, Cefotetan, Pirazinamida, Anfotericina B, Meropenem, Sulfametoxazol, Dapsona, and Ansamicina.

The main search area shows 'Expresión a Buscar:' with the input 'antibiótico', 'Criterio:' with a dropdown set to 'Todos', and an 'ENVIAR CONSULTA' button. Below the search area, it indicates 'Cantidad de Resultados: 71' and 'Ordenar por: Ranking de Búsqueda'. A blue banner highlights the search result: 'Valor de la utilización de espaciadores de cemento y aloinjerto con antibiótico para reconstruir defectos óseos infectados: estudio experimental en conejos'. The details for this result include: 'Autor/es (contribuyente): Bagliardelli, Julio; Diaz Gallardo, Paula; Flores, Jorge; Bustos, Damián; Bruno, Pablo; Allende, Christian', 'Materias: Polimetilmetacrilato, Conejos, Antibiótico, Seudomembranas', 'Editor/Título Revista: Revista de la Asociación Argentina de Ortopedia y Traumatología', and 'Tipo de Material: journal article'. A paperclip icon and text 'Haga click aquí para acceder al Objeto Digital.' are also present. The introduction text at the bottom reads: 'Introducción: El objetivo de este trabajo fue evaluar diferentes formas de tratamiento en fracturas expuestas con pérdida ósea e infectadas en conejos. Materiales y métodos: Se realizó un estudio experimental utilizando 25 conejos adultos. En todos se reprodujo una fractura expuesta mediodiáfisisaria de fémur en la cual se inoculó Staphylococcus aureus. Luego, en forma aleatoria, en el grupo I no se realizó tratamiento (grupo de control), en el grupo II se colocó nilmetilmetacrilato con pentamicina y vancomicina en el foco de fractura asociado a antibiótico por vía sistémica en

Fig. 6 Multiplicación de una consulta a partir de la interfaz experimental de BDU2 (Interfaz experimental de BDU2 <http://cosechador.siu.edu.ar>).

⁴ DBPedia (<http://dbpedia.org>) es la iniciativa que permite extraer información estructurada de Wikipedia y la disponibiliza a través de la web.

Cabe destacar que esta interfaz experimental usa tecnologías emergentes, tales como SPARQL, un lenguaje estandarizado para consulta de grafos RDF, es decir la tecnología que da sustento a la web semántica.

La temática de Repositorios Institucionales y su relación con Linked Open Data posee enormes potenciales para el usuario final, desde habilitar la funcionalidad que permite interconectar el contenido con su autor, conectar el contenido con datos primarios insumo de la investigación tanto como relacionar conceptos, multiplicar las consultas y, en general, facilitar la transición de una simple interfaz de consulta hacia una herramienta de descubrimiento de conocimiento.

Adicionalmente, esta línea de investigación intenta generar resultados en lo relacionado a la interoperabilidad semántica de los repositorios institucionales.

Convenio SIU – Ministerio de Ciencia, Tecnología e Innovación Productiva (MINCyT)

A partir de la creación del Sistema Nacional de Repositorios Digitales y de la estrecha relación entre las instituciones, el MINCyT firma un convenio con el SIU, a través del cual este último le cede en forma gratuita el software que permite la cosecha de los datos a través de la BDU2, para que realice las adaptaciones necesarias para su aplicación para el Sistema Nacional de Repositorios Digitales (SNRD).

Desde la formación del SNRD y la activa participación de los principales repositorios, se descubre que los requerimientos funcionales del software se orientan hacia algunas funcionalidades abarcadas por BDU2, tales como los procesos de cosecha o la presentación de un portal público, y otras funcionalidades relacionadas con la gestión, validación y normalización de los metadatos que los repositorios exponen, entre otras, no consideradas originalmente en el desarrollo de BDU2.

El SIU conforma un equipo técnico entre sus áreas con la finalidad de analizar antecedentes y soluciones similares, teniendo en cuenta la necesidad de dejar capacidad instalada en el Ministerio en relación a la solución, sin dependencia de productos de base comerciales y con licenciamiento apropiado de software libre. Analizadas diversas soluciones que contemplan funcionalidades relevantes por separado, es que se determina la existencia de una solución integral que cumplimenta una buena parte de las necesidades expresadas y que posee una arquitectura actual, tecnología conocida y un grupo de desarrollo activo.

Dicha solución se llama D-NET y es el producto colaborativo del: Istituto di Scienza e Tecnologia dell'Informazione, National Research Council, de Italia, ICM Research Centre Warsaw, de Polonia, National and Kapodistrian University of Athens, Grecia y la Biblioteca de la University of Bielefeld, en Alemania con financiación del proyecto DRIVER de la Unión Europea y actualmente del proyecto Open Aire Plus. El producto se halla bajo licencia Apache.

La Figura 7 muestra un esquema de la arquitectura de D-NET:

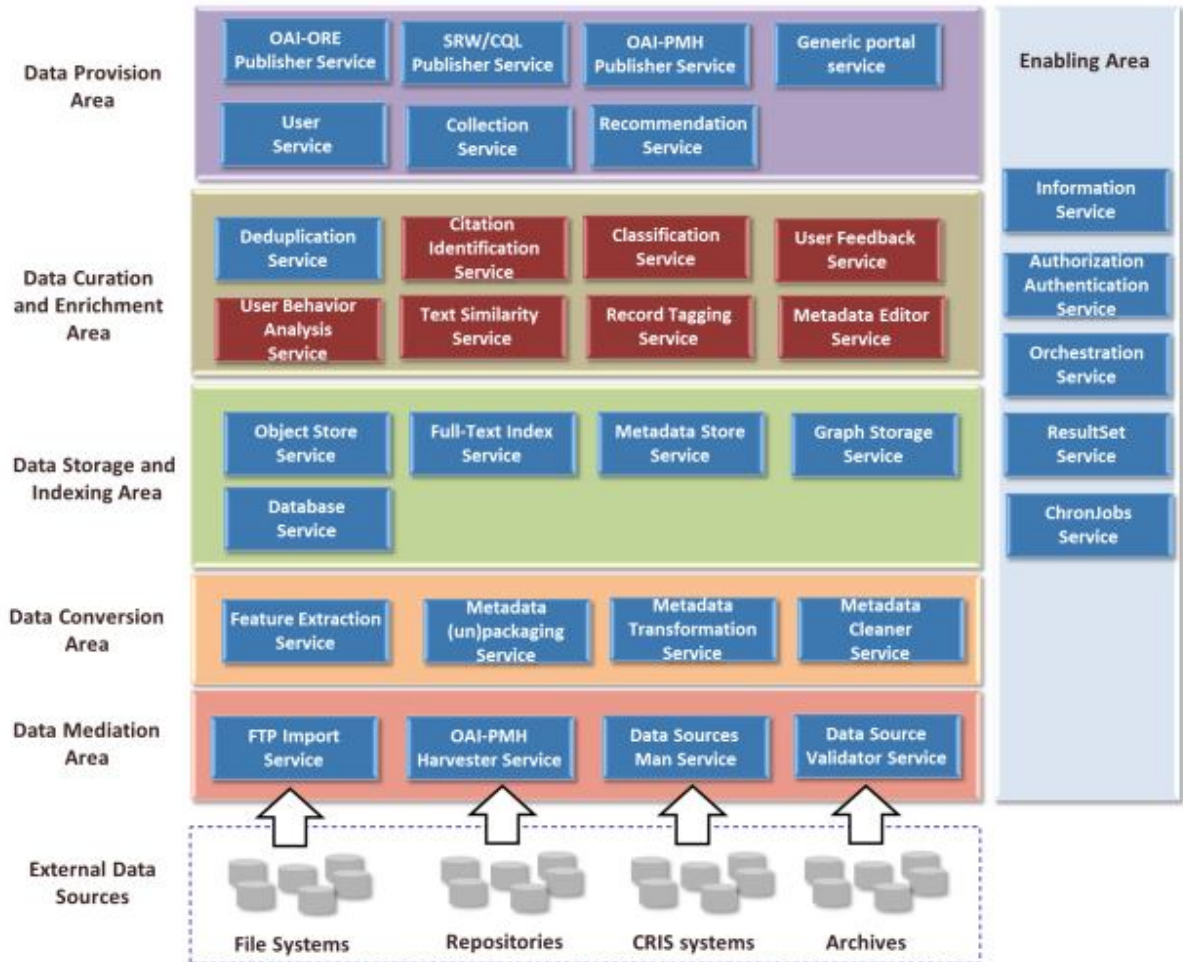


Fig. 7 Arquitectura de DNET mostrando las diferentes áreas y servicios implementados

En la tabla que sigue se describen algunas características de la arquitectura de D-NET y en cursiva un muy breve resumen de las observaciones del equipo de evaluación técnica del SIU.

Característica	Comentarios
<p>Infraestructuras de información. La solución oferta un conjunto de servicios para la colecta de de contenido digital (metadatos u objetos digitales) provenientes de una colección de fuentes de datos para ser provistos a usuarios finales o re-expuestos.</p>	<p><i>El área de provisión de datos implementa uno de los requerimientos expresados por el MINCyT relacionados con la re-exposición de datos (y su potencial conversión) dado que el SNRD es participe de otras redes internacionales.</i></p>
<p>SOA. Arquitectura orientada a servicios. D-NET habilita un framework orientado a servicios donde las aplicaciones consisten en un conjunto de servicios, corriendo potencialmente en nodos de red diferentes. Estos nodos interactúan entre sí para proveer las funcionalidades esperadas por los usuarios.</p>	<p><i>El equipo técnico concluyó en que la arquitectura de la solución seleccionada respeta patrones de construcción de aplicaciones comunes SOA.</i></p>
<p>Customización. Los servicios de D-NET son de propósito general, fueron diseñados para implementar patrones de funcionalidad con lo que pueden resultar fácilmente customizados y re-usados para implementar necesidades específicas de la aplicación en un escenario en particular. Por ejemplo, los servicios de indexación pueden ser configurados dinámicamente para resultar usados con cualquier esquema de metadatos.</p>	<p><i>Esta característica del producto, se ha considerado esencial: El debate sobre el esquema de metadatos que la red del SNRD debe usar se presenta en forma constante. El esquema de metadatos Dublin Core, es un esquema sencillo de aplicar, con múltiples perfiles que permiten complementar su uso, pero presenta limitantes en áreas muy sensibles. Un caso simple es el registro de la afiliación institucional de los autores. Dicha limitante, genera problemas para el uso de los metadatos en un marco de análisis bibliométrico o cuantitativo de la producción volcada en los repositorios.</i></p>
<p>Escalabilidad y automatización. Las aplicaciones de D-NET se hallan formadas por un conjunto de servicios interactuando. Los workflows entre esos servicios pueden resultar orquestados y monitoreados por el Framework.</p>	<p><i>Se ha considerado positivo este aspecto de la solución en vistas a considerar la escalabilidad de la solución en el marco de constante crecimiento de los repositorios institucionales en la región</i></p>

	<i>en general como en Argentina en particular.</i>
Uso de estándares. D-NET adopta el Java Webservice Framework, donde los servicios intercambian datos a través de SOAP, xml es la lengua franca para todo el sistema. Los mecanismos de seguridad son provistos a través del XACML (Access control markup language), implementando lógicamente OAI-PMH y consultas a través del protocolo SRW/CQL.	<i>Se ha considerado positivo el respeto por estándares tecnológicos, más aún considerando que existe experticia en el marco del área de Sistemas del Ministerio de Ciencia y Tecnología sobre estas tecnologías, siendo que el SIU puede acompañar en la transferencia y apropiación de la plataforma por parte de su destinatario.</i>

Es importante destacar que el área de conversión de datos permite continuar con los trabajos relacionados con la Web semántica y su conexión con los repositorios institucionales, permitiendo re-definir la implementación tecnológica de la misma hacia uno o más servicios que pudieran consumir o relacionar los datos obtenidos de los Repositorios Institucionales con datasets Linked Open Data.

Resultados obtenidos en la implementación de D-NET.

El SIU ha mantenido comunicaciones con el equipo de desarrollo de D-NET, ha provisto de una infraestructura tecnológica para su implementación y contando con la colaboración de diversos actores pertenecientes al equipo de desarrollo de la solución, se logró realizar un primer deploy de la misma la cual se halla funcional y trabajando sobre un total de 8 repositorios nacionales a la fecha de cierre del presente artículo. Dicha implementación permitió tomar conocimiento más profundo de la misma, relevar las necesidades técnicas y los aspectos funcionales que sería conveniente, de acuerdo al conjunto de requerimientos locales, reforzar o mejorar.

La interfaz de acceso público se halla customizada (preliminarmente) con la gráfica del Ministerio de Ciencia y Tecnología y se están comenzando las tareas de estudio de los aspectos internos de cada área y cada servicio. Dicho estudio será complementado con una actividad de entrenamiento presencial.

La Figura 8, muestra la interfaz D-NET implementada sobre la infraestructura provista.

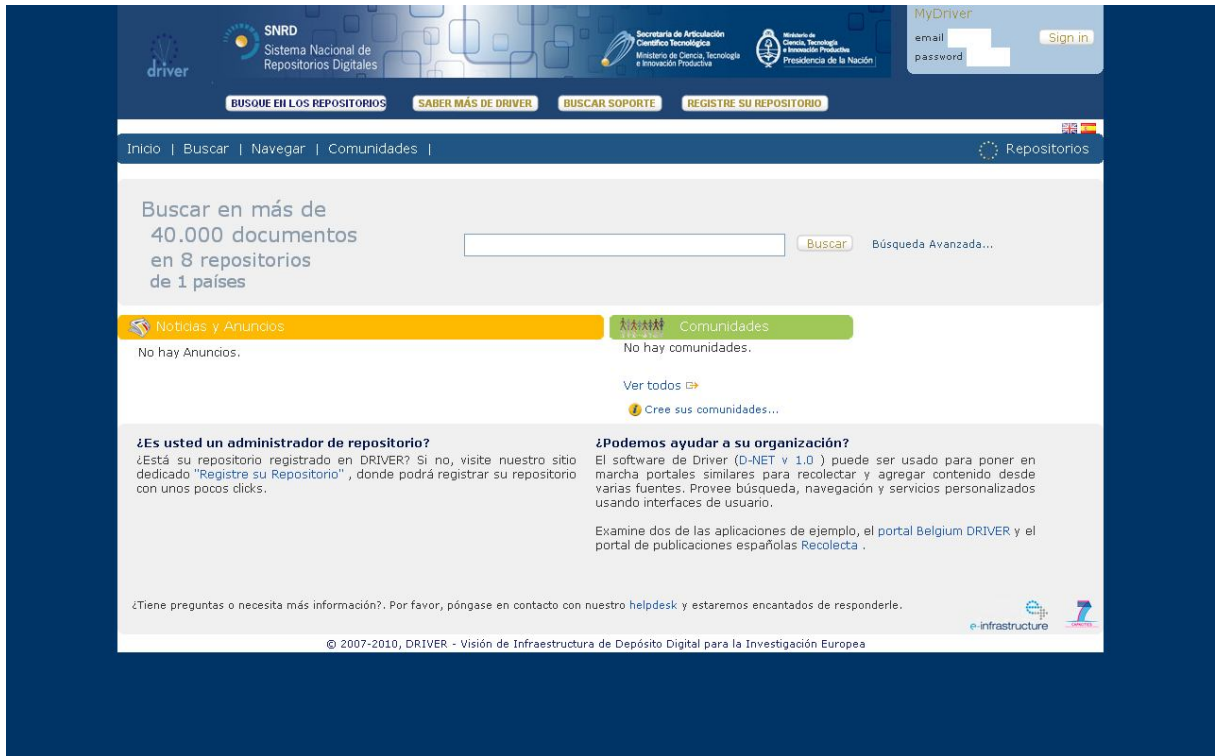


Fig. 8 Interfaz D-NET para el futuro cosechador del SNRD.

Open Science. Hacia donde se mueve la producción del conocimiento científico.

Castelli [11] indica que “...La tecnología ha facilitado que la comunicación y publicación científica consista a día de hoy en mucho más que la traslación digital de la tradicional publicación en papel... Los objetos de información están compuestos de múltiples partes, organizados de acuerdo a estructuras semánticas distintas. Cada parte puede ser de diferente tipo (texto, video, objetos tridimensionales, series temporales, etc)...” continúa indicando que la disponibilidad de estos objetos de información abren nuevas vías para la comunicación científica mucho más ricos y efectivos que en el pasado. Esto impactará claramente en la enseñanza y en la “performance” de la investigación.

“...En este futuro, los investigadores podrán compartir fácilmente información, analizarla, procesarla y validar los resultados por medio de la colaboración de sus colegas y generar así nuevo conocimiento que será inmediatamente volcado nuevamente a la comunidad científica...” Castelli [11]

A su vez Kroes [12] indica "...Comenzamos la era de la Ciencia Abierta....Tomemos por ejemplo el análisis de datos. Análisis de una cantidad grande de datos: Cada año la comunidad científica produce datos que cabrían en el tamaño equivalente a 20 bibliotecas del Congreso de los Estados Unidos...Gran cantidad de datos necesita gran colaboración. Sin ella es imposible coleccionar, combinar y concluir resultados de los distintos experimentos en diferentes países, en diferentes disciplinas.." Cita como ejemplo la secuencia del genoma humano, las bases de datos como EMBL y USGenBank se duplican en tamaño cada 9 meses, sobre una base de 400 billones de entradas de ADN..."

El proyecto Open Aire [13] indica que el desarrollo de e-Infraestructuras para la e-Ciencia es parte de la agenda digital como iniciativa principal, dedicada a conectar investigadores, instrumentos, datos y recursos de computación en Europa. Estos esfuerzos se espera que creen un territorio de "conocimiento abierto" o un "área de Investigación digital europea" donde el conocimiento y la tecnología se mueven libremente. Indica que es esencial tomar una aproximación global para promover la interoperabilidad, descubrimiento del conocimiento y acceso mutuo a los recursos de investigación científica.

Desde el SIU creemos que la tendencia a la creación de objetos digitales complejos, la multiplicación de repositorios específicos de datos tales como Dryad [14], la generación de conocimiento colaborativo en la comunidad científica serán tendencias que se afianzarán en los próximos años. En este marco, siendo que los retos tecnológicos serán crecientes, resulta coherente el hecho de volcar esfuerzos a la apropiación de conocimientos relacionados con la plataforma D-NET, la cual en sintonía con el proyecto Open Aire plus, será la plataforma utilizada (en la presente o en siguientes versiones) para llevar adelante algunas de las funcionalidades necesarias en las citadas infraestructuras de publicación e intercambio de información científica.

Conclusión.

El SIU a través del Módulo Bibliotecas, ha acompañado en los temas de relevancia actuales a las Bibliotecas Universitarias, actuando como un facilitador técnico, proveyendo productos tecnológicos concretos, capacitación y espacios de comunicación y debate para que las mismas puedan abordar tareas que representan desafíos para los cuales en muchos casos, no cuentan con recursos humanos o tecnológicos locales suficientes.

En el caso de los Repositorios Institucionales, se ha trabajado desde el año 2004 ofreciendo la primera capacitación sobre el tema y en el 2009 presentando las ventajas de la adopción de estándares, de la normalización de los datos, del control de calidad de los mismos, de la interoperabilidad sintáctica y semántica. En el marco de dicha tarea es que se crearon los productos como BDU2, los eventos de capacitación y experimentos sobre la web semántica.

Como corolario de dichas realizaciones, otras instituciones de orden nacional, reconocen estos esfuerzos y nos convocan a la labor de profundizarlas, exigiendo tareas de estudio y de investigación hasta alcanzar resultados óptimos. Esto resulta posible dado el grado de sinergia interno del staff SIU: Las implementaciones han sido posibles por el grado de experticia repartido entre áreas internas tales como IT,

especialistas en bases de datos, especialistas en arquitectura de software, y el rol del Módulo como especialista en la temática específica.

Por último es importante destacar la propuesta de fortalecer un proyecto de software libre pre-existente aportando desde documentación hasta modificaciones funcionales y adherir completamente a la cultura de compartir conocimiento, piedra fundacional del acceso abierto en el conocimiento científico

Referencias.

- [1] Suber, Peter (2006). Open Access Overview. Consultado el 11/04/2012. <http://www.earlham.edu/~peters/fos/overview.htm>
- [2] Declaración de Budapest. Consultado el 11/04/2012. <http://www.soros.org/openaccess/translations/spanish-translation>
- [3] Declaración de Bethesda. Consultada el 11/04/2012. <http://www.earlham.edu/~peters>
- [4] Declaración de Berlín. Consultada el 11/04/2012. <http://www.zim.mpg.de/openaccess-berlin>
- [5] Swan, Alma con prólogo de Jānis Kārklīņš. Policy Guidelines for the development and promotion of Open Access. Paris: UNESCO. (2012). Consultado el 10/04/2012. <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>
- [6] Lima Leite, Fernando César. Como gerenciar e ampliar visibilidade de informacao científica brasileira: repositórios institucionais de acesso aberto. Brasilia: IBICT. (2009)
- [7] Gómez Dueñas, Laureano Felipe. Modelos de interoperabilidad en bibliotecas digitales y repositorios documentales: Caso Biblioteca Digital Colombiana – BDCOL. Colombia. (2009) Consultado el 11/04/2012. <http://eprints.rclis.org/handle/10760/14878>
- [8] García Martínez, Ana María. Definición y estilo de los objetos de información digitales y metadatos para la descripción. Boletín de la asociación Andaluza de bibliotecarios, N. 63, Junio 2001, p.23-47
- [9] Miguel, S., Gómez, N. D., Bongiovani, P. Acceso abierto real y potencial a la producción científica de un país. El caso argentino”. EPI, v. 21, n. 2 p.146-153. (2012)
- [10] Gurmendi, Piñero, Marmonti. Web Of data y los repositorios institucionales. Semana de Acceso Abierto 2011. CAICYT-CONICET-MINCyT-OPS/OMS. (2011) http://accesoabiertoargentina.caicyt.gov.ar/files/biblioteca/Ponencia_Pineiro-Marmonti-Gurmendi.pdf
- [11] Castelli, Donatella. Optimising Research Sharing in the European Research Area: Cyberinfrastructure, Quality and Open Access from the Technological Perspective. Rome: European Research Area. (2009). Consultado el 11/04/2011 http://ec.europa.eu/research/conferences/2009/era2009/speakers/papers/paper_donatella_castelli.pdf

[12] Kroes, Neelie. "Open Infrastructures for Open Science". Through e-Infrastructures European Federation of Academies of Sciences and Humanities Annual Meeting. Rome. (2012). Consultado el 11/04/2012.

<http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258&format=HTML&aged=0&language=EN&guiLanguage=en>

[13] Proyecto Open Aire de la Comunidad Económica Europea. Consultado el 11/04/2012.

<http://www.openaire.eu/>

[14] Repositorio de datos Dryad. "...Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a consortium of journals that collaboratively promote data archiving and ensure the sustainability of the repository...". Consultado el 11/04/2012.

<http://datadryad.org/>