

Tecnologías para propagar los contenidos de una Biblioteca Digital. Experiencias desde SeDiCI

Prof. Ing. Marisa Raquel De Giusti (marisa.degiusti@ing.unlp.edu.ar)

Comisión de Investigaciones Científicas de la Provincia de Buenos Aires – CIC

PrEBi-SeDiCI, Universidad Nacional de La Plata, Argentina

A.C. Emiliano Marmonti (emarmonti@sedici.unlp.edu.ar)

A.C. Ariel Sobrado (asobrado@sedici.unlp.edu.ar)

A.C. María Marta Vila (vilamm@sedici.unlp.edu.ar)

A.C. Gonzalo Luján Villarreal (gonetil@sedici.unlp.edu.ar)

PrEBi-SeDiCI, Universidad Nacional de La Plata, Argentina

Resumen:

En el panorama actual de las Bibliotecas Digitales puede observarse la creciente importancia de admitir protocolos que permitan la provisión de servicios a otras interfaces que accedan directamente, con las consideraciones apropiadas de seguridad, oportunidad y pertinencia, a los repositorios administrados por la Biblioteca Digital.

Dichas interfaces pueden ser, por ejemplo, clientes RSS para novedades en los contenidos o un medio para Diseminación Selectiva, metabuscadores, OAI Service Providers, portales web internos de las mismas unidades académicas que acceden a la biblioteca digital para ofertar a sus propios usuarios información residente en los repositorios, o incluso plataformas de e-learning que precisen integrar la información administrada como parte de los contenidos o como bibliografía a consultar.

Desde el año 2004, el Servicio de Difusión de la Creación Intelectual (SeDiCI-UNLP), brinda la posibilidad de devolver parte de sus acervos electrónicos a través de webservices basados en la tecnología SOAP. Sin embargo, dicha implementación no se halla basada en ningún estándar para acceder y consultar los recursos digitales administrados, otras plataformas tales como Fedora, por su misma arquitectura admiten naturalmente una variedad de "conexiones" al repositorio de objetos digitales administrados.

SeDiCI a su vez, es OAI Data y Service provider, siendo que buena parte del material colectado puede ser de utilidad a las bibliotecas al momento que los usuarios realizan consultas en sus respectivos catálogos locales en la UNLP. En un universo de usuarios que acceden a Google como fuente primaria de información, las Bibliotecas intentan día a día "competir" contra los metabuscadores y lograr que sus usuarios accedan al catálogo local como primer opción, SeDiCI tiene la responsabilidad de proveerles en forma transparente y sumamente performante ofertas de material accesible libremente y en línea, tanto proveniente del acervo local administrado como de la información referencial colectada a través del protocolo OAI. Una solución genérica para la atención tanto de los requerimientos de OPAC's internos como de interfaces externas a la UNLP puede ser implementada mediante la adición de una interfaz de consulta a los acervos electrónicos de SeDiCI, a través del estándar SRU/W. Dicho estándar, que replica las características de avanzada del protocolo Z39.50 en cuanto a acceso y facilidades de búsqueda, está actualmente siendo adoptado por diversas plataformas de Bibliotecas Digitales (como es el caso de DSpace).

El objetivo de esta ponencia será presentar un estudio de avance junto con los resultados, así como también el impacto de la implementación de SRU/W sobre el repositorio de documentos locales de SeDiCI. También se estudiará el grado de complejidad basado en la bibliografía existente acerca de la implementación de un gateway SRU/W que permita acceder a la información referencial obtenida a través de OAI y el grado de impacto esperado en los OPAC's de la Universidad Nacional de La Plata.

La idea rectora de esta clase de servicios es la de fortalecer los "puntos simples de acceso" y reducir la cantidad de sitios a los que los usuarios deben acceder para hallar información académica generada en sus propias instituciones. En el marco de dicha tendencia, se debatirá si la estrategia en software de administración de Bibliotecas Digitales apunta a simplificar o incluso desaparecer la interfaz web autónoma y transformarse en un conjunto de servicios-web que son usados desde los diversos portales internos o externos en forma transparente.

Palabras clave: OAI; SRU/W; Z39.50; webservices

Sección 1.

Protocolo Open Archives. Repositorios Dinámico y Estático.

1. Protocolo OAI. Importancia y Difusión.

La Iniciativa Archivos Abiertos (Open Archives Initiative - OAI) nace con el objetivo de desarrollar estándares de interoperabilidad para la difusión de contenidos en Internet, permitiendo de este modo el acceso a un gran número de materiales digitales y mejorando el acceso a las publicaciones electrónicas

OAI es un protocolo de intercambio de información referencial basado en XML acerca de "documentos" existentes, en muchos casos hace referencia a recursos web disponibles libremente, ejemplo: tesis en línea, informes, reportes, artículos, imágenes, etc. En la iniciativa OAI participan Universidades e instituciones de todo el mundo, incluyendo EEUU, Europa, Brasil, etcétera, y abarca las más variadas temáticas.

Los roles definidos por el protocolo pueden ser los de proveedores de datos (Data Providers) y/o proveedores de servicios (Service Providers). Un servidor que se dice Data Provider mantiene uno o más repositorios que soportan el protocolo OAI-PMH como un método para exponer metadatos; un Service Provider es aquel que realiza los requerimientos OAI-PMH hacia los Data Providers, obteniendo metadatos de ellos y utilizando esta información como una base para construir u ofrecer servicios con valor agregado.

2. Implementación del rol de Service Provider en SeDiCI

En SeDiCI se implementó un software de recolección capaz de conectarse con servidores de datos OAI y obtener la información referencial disponible; dicho software está basado en la librería "harvester" de OCLC [OCLC2]. El material así obtenidose ofrece a los usuarios, los cuales pueden realizar búsquedas sobre los documentos alojados en el servidor, y si lo desean, en la información obtenida por medio del harvesting (opción Búsquedas, repositorios externos).

El software tiene predefinidos una serie de repositorios externos sobre los que realizará el harvesting, la selección de los repositorios no es aleatoria, sino atendiendo a las áreas requeridas por los usuarios. La selección de los destinos es dinámica y se basa en el feedback y los índices de uso de cada repositorio.

3. Implementación del rol de Data Provider en SeDiCI. Repositorio Estático.

SeDiCI es también en la actualidad, Data Provider debido a lo cual toda la información y documentos alojados localmente pueden ser accedidos por medio del protocolo OAI-PMH. El rol de Data Provider requiere de un archivo XML para exponer los documentos y permitir que la información sea accedida y recolectada por algún proceso harvester (recolector). En las primeras experiencias realizadas el archivo XML se generaba estáticamente, lo que permitía un método rápido y sencillo para exponer el acervo documental de SeDiCI. A posteriori fue incorporada la herramienta YAR [YAR] que genera una versión comprimida del XML, obteniendo archivos en un formato conocido como XMLTape. Este formato puede ser usado para almacenar miles de registros XML de manera muy eficiente, en un archivo con formato ampliamente conocido y soportado, mientras se provee una interfase sencilla OAI-PMH sin la necesidad de una compleja administración.

Las facilidades provistas por esta herramienta permitieron transformar a SeDiCI en Data Provider en relativamente poco tiempo y con costos de implementación relativamente bajos; junto con la ventaja de realizar una mejor utilización de los recursos propios al almacenar la información en un formato comprimido.

Pero el constante ingreso de nuevos documentos implicaba que el archivo XML que los expone debía ser generado constantemente, con lo cual se debía también generar nuevamente el archivo comprimido con formato XMLTape. Si bien esta tarea se realizaba en forma programada y automática, la herramienta yar hasta la versión utilizada no implementaba la división en Sets y desde la perspectiva del desarrollo resultaba más ordenada la generación de la información OAI expuesta utilizando de un script dinámico.

4. Implementación del rol de Data Provider en SeDiCI. Acceso dinámico. (OCLC - software libre -)

Las características propias de SeDiCI, que dentro de su mandato incluye el uso de recursos tecnológicos para maximizar la difusión de la información contribuida, obligó la búsqueda de una solución que implemente en su totalidad el protocolo Open Archives, admitiendo todas las facilidades que el protocolo define. Para tal fin, se ha implementado en SeDiCI un nuevo Data Provider, basado en una librería desarrollada por la Universidad de Oldemburgo en Alemania [OLD] que se denomina PhpOAI2.

Sección 2.

Beneficios de exponer Metadatos. Responsabilidad del Servicio de Difusión de la Creación Intelectual (SeDiCI).

1. Responsabilidades de SeDiCI en la Difusión de los trabajos académicos que los usuarios contribuyen

Unos de los objetivos con el que se realizó SeDiCI es el de crear un punto de contacto entre los autores de los trabajos, investigadores, desarrolladores y grupos de trabajo, y las personas que acceden a sus trabajos a través de alguno de los métodos de consulta existentes; para ello SeDiCI permite incorporar toda la información de los creadores de cada trabajo junto con la información del trabajo en sí, siempre que cada persona consienta que se muestre esta información (caso contrario, la información permanecerá oculta para el público en general). Esto permite a los autores de los trabajos entrar en contacto con investigadores de temas afines en otras partes del mundo, vinculándose con otros círculos científicos de distinta complejidad y principalmente, evitando la duplicación de esfuerzos y permitiendo mejorar el desarrollo de temas de investigación localmente.

2. Tipos de materiales que administra SeDiCI. Relevancia Académica de los mismos.

La Universidad Nacional de La Plata posee actualmente 15 unidades académicas, y en cada una de ellas se realiza un sin número de desarrollos, que abarcan no solo una gran cantidad de áreas sino también una gran variedad de formatos, gracias a la heterogeneidad que caracteriza a dicha Casa de Altos estudios.

La diversidad de formatos de los documentos presentó la necesidad de desarrollar una plataforma capaz de mostrar cualquier trabajo sin importar su formato o la disciplina a la que pertenece. El software Celsius DL permitió incorporar desde documentos de texto hasta cuadros o incluso documentos musicales.

Dentro de la UNLP se generan también un gran número de revistas científicas, y esta información también debe poder ser vista y accedida desde SeDiCI. Para ello, la plataforma de software permite acceder de manera ordenada a estas revistas y a las partes que las componen, entre las que se incluyen la tapa, portada, contraportada, índice, editorial, reseña, y los documentos en sí, todo esto

para cada volumen de las revistas.

SeDiCI funciona también como editorial de libros electrónicos, permitiendo la generación y publicación de eBooks dentro del portal. Este tipo de material también puede ser visto de manera organizada, accediendo organizadamente a cada una de las partes que lo componen (portada, índice, capítulos...)

Sección 3.

Experiencias previas de webservices como interfaz de acceso. Catálogos en línea que acceden a los webservices de SeDiCI. Experiencias futuras en el uso de webservices. Casos de estudio.

El uso de webservices ha crecido mucho en los últimos años y continúa haciéndolo a medida que se descubren las ventajas y posibilidades que ofrecen. Los webservices permiten que dos equipos se comuniquen entre sí por medio de un protocolo llamado SOAP, relativamente simple, eficiente y que permite abstraerse no solo de la plataforma en sí (Sistema Operativo, motor de Base de Datos) sino incluso del lenguaje de programación sobre el que se trabaje, gracias a la utilización del formato XML para la transferencia de información. También permite una gran heterogeneidad en cuanto al tipo de información que se maneja de cada lado de la comunicación, gracias a la utilización de archivos de configuración (archivo WSDL, también con formato XML) que permiten indicar explícitamente con que información se cuenta, que datos se presentan y de que tipo es cada uno. Este archivo define todos los servicios que cada interfaz de acceso de webservices proveerá.

En el área de Bibliotecas Digitales, la tecnología de webservices está siendo muy utilizada, principalmente para la intercomunicación entre Catálogos en línea, permitiendo así aumentar la oferta de información al usuario final.

1. Caso de estudio. Catálogo de la Biblioteca de Física de la UNLP. Integración de Tecnologías.

En SeDiCI se ha implementado una interfaz de acceso al acervo bibliográfico por medio de webservices; los catálogos online pueden realizar consultas a SeDiCI por medio de esta tecnología y de este modo, ofrecer a sus visitantes información sobre tanto local como información contenida en SeDiCI.

Gracias a la utilización de webservices, el OPAC de la Biblioteca de Física de la UNLP (basado en el software Koha-BDF [Koh]) se comunica con SeDiCI y obtiene de allí todos los documentos que pertenecen al esa área que están almacenados en SeDiCI. Los resultados pueden ser explorados por separado o incluidos en los resultados referenciales del acervo propio de la biblioteca.

La portada del catálogo en línea de la Biblioteca de Física [FIS] presenta información estadística sobre los documentos alojados en SeDiCI que son accesibles desde allí. Los usuarios de la Biblioteca también pueden realizar búsquedas sobre estos documentos sin la necesidad de salir de dicho portal ya que, por medio de los webservices, toda la información que SeDiCI posee sobre dichos documentos es "entregada" para que el portal de Física le de el formato que necesita y la muestre de acuerdo a sus estándares propios, y de manera totalmente transparente para el usuario.

Por último, el portal de Física también brinda un enlace que llevará al usuario al documento en portal de SeDiCI, desde donde podrá disponer del documento y de todas las partes que lo componen de manera inmediata.

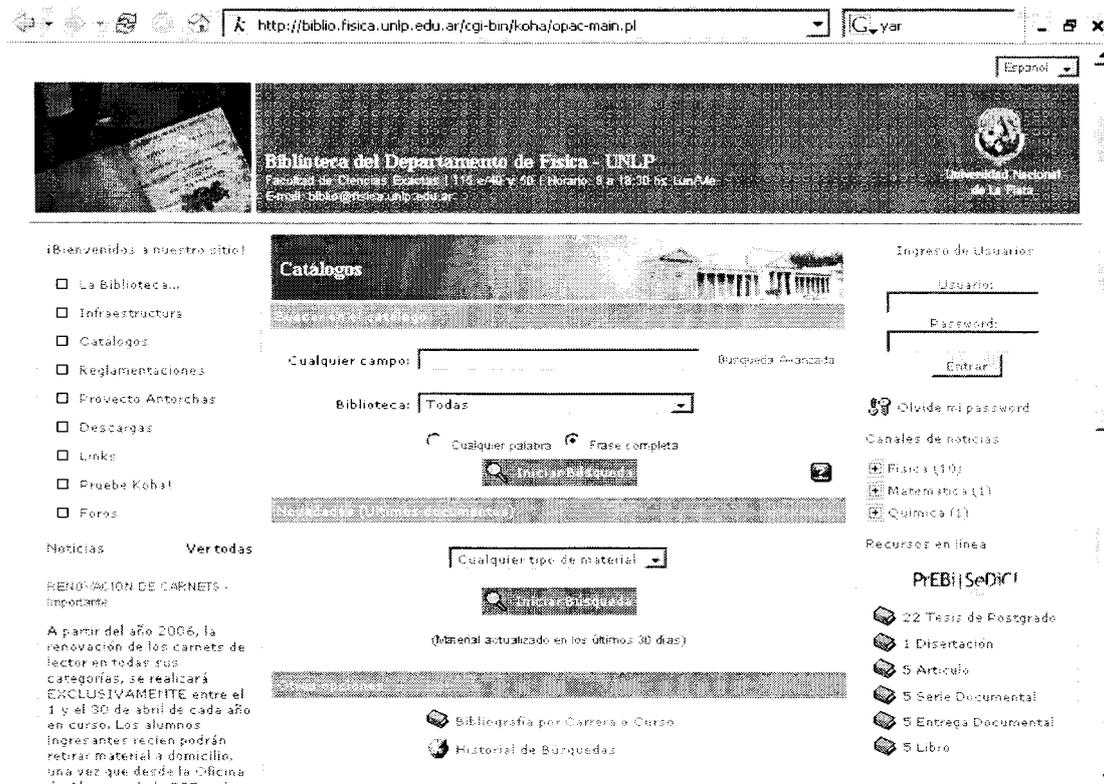


Figura 1. El OPAC de BDF "consume" webservices provistos por la Biblioteca Digital SeDiCI.

2. Otros casos de uso en el uso de webservices entre software para Bibliotecas.

La plataforma Celsius [CEL] para gestión de pedidos del Proyecto de Enlace de Bibliotecas (PREBi) está evolucionando hacia la tecnología de webservices, lo cual es la base de Celsius Network. Esto permitirá que las distintas instancias de Celsius instaladas en toda América y España permitan comunicarse entre sí, para finalmente lograr una automatización en muchas tareas que hasta ahora requieren la intervención directa de los operadores. Por ejemplo, Celsius permitirá generar remotamente un pedido a partir de la información local; esto se realizará enviando por medio de webservices toda la información del pedido hacia la instancia correspondiente de Celsius, junto con información sobre la operación que se desee realizar. El movimiento de documentos entre los distintos Celsius también podrá realizarse mediante el uso de webservices, ya que los mismos permiten mover de manera segura, ya no solo información sobre un documento específico, sino el documento en sí (embeber el objeto documento en el envelope del SOAP).

La interconexión de instancias de Celsius presenta varios puntos críticos en materia de seguridad, sobre los que se deberán acentuar los esfuerzos; por ejemplo, el intercambio de información sensible (datos de los usuarios, documentos y eventos de pedidos) implicará la encriptación de los datos en cada Celsius previo a su envío.

En lo relativo a la plataforma utilizada, las tecnologías subyacentes debajo de los protocolos resultan por completo indiferentes, gracias a que el protocolo garantiza un medio de comunicación neutro y estable.

Se ha puesto en marcha otro proyecto en el marco la UNLP que permitirá conectar dos sistemas de gestión de bibliotecas, uno basado en tecnologías Microsoft y el otro basado en tecnologías OSS (Koha-BDF) para el intercambio de información relacionada con acervos, con estados de cuenta del usuario y con transacciones de los mismos.

Desde el punto de vista del usuario implica un sistema unificado virtual basado casi en su totalidad en webservices, sin obligar a migrar las plataformas de sistemas de gestión de ambas instituciones ni los datos que las mismas administran.

Sección 4.

El protocolo Z39.50. Análisis. Ventajas y Desventajas.

1. El protocolo Z39.50. Breve descripción técnica del protocolo.

El protocolo Z39.50 es un estándar internacional para la comunicación entre sistemas informáticos, y es cada vez más importante en el desarrollo de sistemas en el ámbito de las Bibliotecas Digitales. Es un protocolo basado en la filosofía Cliente – Servidor, es decir, existen clientes que hacen requerimientos (request) hacia uno o más servidores, y estos reciben los requerimientos, los procesan, y envían al cliente el resultado de dicho procesamiento. Z39.50 no solo provee un método eficiente para recuperar información, sino que también facilita la interconexión de sistemas informáticos, sin importar el software y hardware que corran los sistemas involucrados.

• El protocolo Z provee los siguientes servicios:

○ Posee 11 servicios nativos:

- Init, Search, Present, DeleteResultSet, AccessControl, ResourceControl, TriggerResourceControl, ResourceReport, Scan, Sort, Segment

○ Posee 7 servicios extendidos:

- PersistentResultSet, PersistentQuery, PeriodicQuerySchedule, ItemOrder, DatabaseUpdate, ExportSpecification, ExportInvocation

2. Casos de uso en Argentina. Problemas y Soluciones.

En Argentina, la difusión del formato MARC21 es relativamente incipiente, existen iniciativas de conversión retrospectiva desde formatos bibliográficos anteriores, las cuales fueron concretadas y en algunos casos, en línea a través de los correspondientes OPACs. El incentivo principal se basa en dos pilares: el uso de la catalogación por copia, y la posibilidad de compartir la información referencial acerca de los acervos de producción propia que las bibliotecas poseen. El Ministerio de Educación de la Nación, a través del Programa SIU y con sustento económico de Fundaciones locales ha realizado un importante esfuerzo en lo relacionado a conversión retrospectiva y capacitación en el formato, inclusive con programas de capacitación a distancia [SIU].

El uso del protocolo Z39.50 de momento se encuentra limitado al uso bajo el perfil de cliente. Un hecho decisivo en este sentido ha sido la difusión de la herramienta UNESCO-IsisMarc [ISM], la cual permite catalogar registros MARC21 usando la tecnología Isis y a su vez, gracias a la biblioteca OSS Yaz [YAZ], incluye un cliente Z39.50 para acceder a diversos repositorios desde los cuales descargar registros, como por ejemplo Library Of Congress u Ohio Link por citar los mas masivos de ellos.

Una dificultad hallada en el uso del protocolo, consiste en lograr que las Unidades Informáticas de las Instituciones académicas habiliten el uso del amplio abanico de puertos TCP/UDP de las máquinas destinadas al proceso de catalogación. Si bien existen iniciativas para normalizar el uso del puerto 210 para los servidores Z, la misma Library Of Congress continúa usando el puerto 7090. Esto presenta una dificultad a la hora de convencer a los administradores de redes acerca de las necesidades técnicas de los puestos de catalogación. En vista de la masa crítica de uso del protocolo, y los potenciales beneficios que el uso del mismo puede brindar, es que actualmente se está estudiando una solución técnica para estos problemas basada en la herramienta Yazproxy de indexdata, la cual al menos paliaría el problema de apertura de puertos desde las múltiples computadoras destinadas a procesos técnicos.

Desde esta perspectiva, actualizar el protocolo Z de acuerdo a lo enunciado en Z39.50 new generation [zin] , poniendo en su capa de transporte el protocolo http, y como "envoltorio" el protocolo XML brindaría una solución a los problemas planteados en el párrafo anterior.

Es importante en este aspecto sensibilizar al personal de gestión de las bibliotecas acerca de la importancia que tiene maximizar por medios tecnológicos la difusión del material de producción propio, no solo para la biblioteca en sí y los potenciales servicios que a partir de dicho intercambio pueden generarse (canje, referencia, préstamo interbibliotecario, etc), sino para los autores del material objeto. En ese aspecto, SeDiCI puede dar constancia de los intercambios y contactos producidos a través de la Biblioteca Digital.

3. Redes de Metabúsqueda basadas en el protocolo.

Las necesidades de metabúsquedas y metabuscadores tienen su punto más fuerte en uno de los servicios fundamentales al día de hoy en las bibliotecas: la referencia. La experiencia del Proyecto de Enlace de Bibliotecas (PrEBi-UNLP [PREBI]) en este sentido es única a nivel nacional.

Desde 1997 la UNLP dispone de recursos para generar un servicio de referencia central altamente especializado en la obtención de documentos de apoyo a la gestión de investigación. Este servicio es transversal a todas las disciplinas en la UNLP y atiende a grupos de investigación tanto en ciencias duras como en ciencias humanísticas; mediante este servicio se han provisto un total 20.000 documentos hasta la fecha. Dentro de los recursos que la UNLP destina para el funcionamiento del PrEBi se encuentra la suscripción al consorcio ISTECA, el cual brinda el acceso a alrededor de 40 catálogos en línea donde es posible ubicar potencialmente el material solicitado por los usuarios. Es de destacar que al día de hoy y a pesar de funcionar en forma activa el intercambio de documentos desde 1997, el consorcio no cuenta con un metabuscador de recursos bibliográficos para las bibliotecas cooperantes, siendo ésta una de las más positivas y encomiables líneas de gestión actual de la iniciativa DLL [ISTECA]. Esto incrementa en forma considerable el esfuerzo de un referencista para poder solicitar un documento, debiendo basarse en herramientas auxiliares, principalmente su expertise o la ayuda que un sistema informático pueda brindarle de alguna forma como "ayuda-memoria".

Existen iniciativas de metabuscadores comerciales o desarrollos que combinan el acceso a diferentes tipos de material. Una de las experiencias en cuanto a metabuscadores es Library Of Congress con el acceso a través del Gateway Z adaptado de una pieza de software del CNDIR [LOC] o la red nacional de Bibliotecas de Rusia [RUS]. Otra experiencia importante para relatar es la del Canadian Geospace Knowledge Network [CAN], que desde su página permite descargar el paquete de software que implementa un servidor Z39.50 para cada uno de sus nodos cooperantes, y más aún, no basan su estándar de intercambio en el formato MARC21, sino en el lenguaje de marcado XML y su propio esquema de metadatos (GEO).

Otras iniciativas basan sus metabuscadores en una interpretación de los resultados de los diferentes OPAC's a través de un método de "fuerza bruta" o web scrapping, con el cual si se produce un cambio en la visualización de los resultados de una consulta en el OPAC, el metabuscador perdería la posibilidad de acceder a los resultados del mismo.

En este marco, desde PrEBi-SeDiCI creemos sumamente importante, mas que desarrollar piezas de software, establecer acuerdos en relación a protocolos neutros y si es posible basados en XML para el acceso a cada catálogo, como una interfaz alterna, no destinada al público, sino a las redes que accedan al mismo. Dicho protocolo, debe resultar lo mas estable y mantenible en el tiempo posible.

Las tecnologías subyacentes, si bien son un problema por la variedad de sistemas comerciales o por los modelos de acceso a la información, en realidad representan un obstáculo menor. Sustentamos dicha afirmación en la existencia de herramienta OSS Simple Server Z39.50 de indexdata [SSE]. La misma implementa a través de un módulo del lenguaje Perl los servicios básicos de acceso al protocolo Z39.50 abriendo la implementación de los mismos a cualquier tecnología. En el desarrollo del presente trabajo veremos la experiencia de SeDiCI en el uso del mismo.

Sección 5.

Z39.50 New Generation (ZING). Protocolos SRW/U. Análisis. Ventajas y Desventajas.

1. Evolución del Protocolo. ZING. CQL.

ZING (Z39.50 International Next Generation) es el último proyecto de la Z39.50 International Maintenance Agency. Abarca una serie de iniciativas promovidas por los impulsores de Z39.50, con el fin de hacer más accesibles tanto los contenidos intelectuales como semánticos de dicho protocolo para finalmente transformarlo en una herramienta mas atractiva tanto para desarrolladores y usuarios como para proveedores de información y distribuidores, y sobre todo, ahorrando esfuerzos para ponerlo en funcionamiento.

2. CQL.

(Última versión: 1.1)

CQL es un lenguaje formal para representar consultas para sistemas de recuperación de información como ser índices web, catálogos bibliográficos e información sobre colecciones de museos. La principal premisa con la que se diseño CQL es que las consultas puedan realizarse en un lenguaje "humano", que sea fácil de escribir y de leer, que sea lo más intuitivo posible, y sobre todo esto, que mantenga la expresividad de lenguajes de consulta mas complejos. CQL trata de combinar en las expresiones la simplicidad e intuitividad de los lenguaje más simples (CCL, Google) para realizar las consultas de todos los días, con la riqueza de los lenguajes de consulta más expresivos (SQL, PQF, XQuery) para permitir la aplicación de conceptos más complejos cuando es necesario.

3. Servicios web: SRW y SRU.

SRU (Search / Retrieve Via URL) es un protocolo estándar de búsqueda para realizar consultas a través de Internet. Este protocolo utiliza CQL como lenguaje para realizar dichas consultas.

Una de las principales ventajas de SRU radica en su sencillez. Un requerimiento SRU desde un cliente a un servidor es simplemente una URL HTTP, la cual consta de dos partes: la url base, y la parte de

búsqueda. Los requerimientos SRU pueden incluir caracteres Unicode.

La url base es simplemente la dirección (y puerto) del servidor contra el que se hará la consulta. La parte de búsqueda consiste en una serie de parámetros, separados por el carácter "&", con la forma "clave=valor".

El servidor recibe el requerimiento y parsea la url, obteniendo los nombres de los parámetros y los valores de los mismos. Luego de verificar los nombres de cada parámetro, el servidor ejecuta la consulta especificada y genera una respuesta con formato XML, la cual es enviada al cliente para que la interprete.

La idea de SRW es muy similar a la de SRU. La diferencia radica en que los mensajes entre el cliente y el servidor no son enviados a través de una URL, sino que se utiliza el formato XML sobre el protocolo HTTP, por medio de SOAP. Si bien a primera vista la diferencia parece ser mínima, este cambio acarrea un conjunto de ventajas muy significativas. Los beneficios que se desprenden directamente son: mejor soporte para extensión, autenticación, y características de webservices.

En comparación a Z39.50 puro, una implementación basada en SRU/SRW aportará una claridad mucho mayor, lo cual siempre aporta a la extensibilidad de las aplicaciones. Todo esto con la enorme ventaja de estar siempre apegados a un estándar, lo cual asegura tanto documentación como funcionalidad verificada.

Sección 6.

Caso de Estudio. Implementación de SRU/SRW para el repositorio digital SeDiCI.

1. Desafíos en la integración de diferentes tecnologías.

Como ha sido explicado con anterioridad, en lo incumbente a la adopción de los diferentes protocolos con la finalidad de difundir el material académico hemos abordado una prueba piloto relacionada con el uso de SRW/SRU sobre el repositorio de documentos digitales de SeDiCI.

Nuestro objetivo por tanto, es analizar si dicho protocolo es posible de implementar con las herramientas tecnológicas OSS disponibles, los costos hasta alcanzar el "know how" necesario para implementar la totalidad de servicios considerados en Z39.50 sobre la capa SRU/W, y el costo en esfuerzo de adaptación al repositorio local de metadatos administrado por SeDiCI. Es asimismo uno de los objetivos, analizar los servicios adicionales que podría brindar SeDiCI a las bibliotecas con la implementación del mismo.

Cabe destacar que SeDiCI administra su repositorio de metadatos basado en un motor de base de datos relacional OSS. Si bien se han realizado experiencias en desarrollos a terceros basados en motores XML-native (tales como eXist), la migración del repositorio principal aún no ha sido concretada.

2. Casos de implementaciones previas usando Simple Server.

La herramienta Open Source Simple Server de indexdata cuenta con una vasta experiencia de uso a nivel internacional, tales como el caso relatado en secciones anteriores del Canadian GeoSpace, implementaciones de Servidores Z39.50 para productos basados en tecnología UNESCO/ISIS y asimismo en plataformas de software libre como Koha en su módulo servidor Z39.50. A partir de sus últimas versiones, Simple Server implementa no solo el acceso a través del protocolo Z, sino que

adicionalmente provee implementación de los protocolos SRU y SRW adicionando la librería yaz.

Cabe destacar que estos productos son completamente fiables para un entorno de producción, solo cabe en este punto destacar el caso de uso de Library Of Congress, la cual al momento de implementar SRU/SRW para su servidor Z39.50 utilizó el producto OSS indexdata yazproxy [YAZPROXY] actualmente en producción.

3. Caso de implementación sobre SeDiCI.

El proceso consistió en la instalación del producto Indexdata Simple Server en nuestro servidor y en la implementación de los servicios básicos de acceso y consulta sin considerar de momento consultas complejas como operadores lógicos cualificados sobre diferentes campos o servicios tales como el Scan (propio de los servidores Z).

Las decisiones tomadas consistieron en proveer solo acceso a la información referencial existente en el repositorio principal de producción propia (no en las referencias obtenidas a través de OAI). El schema de registros a devolver como resultado de las consultas se limita a Dublin Core, al menos en esta versión experimental.

Cabe destacar asimismo que Celsius-DL, la plataforma que implementa SeDiCI usa un esquema libre de metadatos, en muchos casos mas rico que Dublin Core, pero con una compatibilidad certificada en la implementación del servicio OAI-Data Provider.

En futuras extensiones es posible (dado que el proyecto original de SeDiCI lo contemplaba) admitir mapeos hacia otros recordSchemas tales como MARC-XML o MODS.

Hallazgos y dificultades:

1. Desarrollo de laboratorio planteado en una duración de 2 (dos) meses. Escalado en dos subobjetivos: Implementación del servidor Z e implementación de los servicios SRU/W.
2. Implementación en el lenguaje de programación Perl: La tendencia de los desarrollos de PreBi-SeDiCI orienta hacia las tecnologías J2EE. Perl es un lenguaje de extrema potencia y amplísimo rango de uso, pero por su "antigüedad", no brindado en la currícula de las carreras de informática. Con parecidos y diferencias con un lenguaje como PHP. Resultado: Curva de capacitación necesaria.
3. Servicios resultantes: SeDiCI pasa a ser un servidor Z39.50 y un "implementador" del servicio SRU/SRW. Contemplar el conjunto de servicios y combinaciones de uso de atributos BIB-1 que oferta el protocolo Z implica un esfuerzo importante de desarrollo y testing.
4. El producto IndexData Simple Server se demuestra realmente maduro, aunque no sucede lo mismo con la documentación que acompaña al producto. Particularmente existe la necesidad de mejorar la documentación relacionada con la implementación y configuración de los protocolo SRU/W (si bien el mismo es la modificación mas reciente del producto).

```

- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2006-04-22T20:11:34Z</responseDate>
  <request verb="ListRecords" from="2004-01-01" until="2005-12-31"
  metadataPrefix="oai_dc">http://sedici.unlp.edu.ar/phpoai/oa2.php</request>
- <ListRecords>
- <record>
- <header>
  <identifier>ARG-AABRA-ART-0000000001</identifier>
  <timestamp>2004-12-28</timestamp>
  <setSpec>6</setSpec>
</header>
- <metadata>
- <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
- <dc:title>
  Variación regional de la tendencia secular de la talla adulta masculina en la provincia de Jujuy
</dc:title>
<dc:title>Revista Argentina de Antropología Biológica</dc:title>
<dc:title>1996, vol. 1, no. 1</dc:title>
<dc:creator>Bejarano, Ignacio</dc:creator>
<dc:creator>Dipierri, José Edgardo</dc:creator>
<dc:creator>Ocampo, Susana Blanca</dc:creator>
<dc:subject>Antropología Biológica</dc:subject>
<dc:subject>Series temporales</dc:subject>
<dc:subject>Altura corporal</dc:subject>
- <dc:description>
  La tendencia secular (TS) de la talla adulta masculina (TESTAM) ha sido estudiada en diversas poblaciones de

```

Figura 2. Implementación de OAI-Data provider en Sedici

```

- <zs:searchRetrieveResponse>
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>8</zs:numberOfRecords>
- <zs:records>
- <zs:record>
  <zs:recordPacking>xml</zs:recordPacking>
- <zs:recordData>
- <record>
- <metadata>
  <Creator>Llorente, Berta Elizabet</Creator>
  <Contributor>Caffini, Néstor O.</Contributor>
  <Contributor>Giulietti, Ana María</Contributor>
- <Title>
  Aislamiento, purificación, caracterización y producción in vitro de peptidasas de alcaucil coagulantes de la
  leche
  </Title>
  <Subject>Bioquímica</Subject>
  <Subject>Péptidos</Subject>
  <Date>2002-01-01</Date>
  <Language>Español</Language>
  </metadata>
</record>
</zs:recordData>
  <zs:recordPosition>1</zs:recordPosition>
</zs:record>
- <zs:record>
  <zs:recordPacking>xml</zs:recordPacking>
- <zs:recordData>

```

Figura 3. Implementación de una consulta a través de SRU sobre el repositorio de sedici (véase la devolución de metadatos bajo el esquema Dublin Core)

En la Figura 3 puede observarse el acceso al repositorio local bajo la expresión de consulta Bioquímica. La URL, SRU para dicho acceso es:

<http://sedici.unlp.edu.ar:8080?version=1.1&operation=searchRetrieve&query=Bioquimica&maximumRecords=10&recordSchema=dc>

Descomponiendo la misma:

version = versión del protocolo

operation= tipo de operación contemplada en SRU/W

query=expresión de consulta

maximumRecords=<x> cantidad de registros a visualizar por búsqueda.

recordSchema=se refiere al esquema XML a devolver. En el caso de SeDiCI solo DC.

De igual forma que para el protocolo OAI es posible utilizar OAI Explorer para determinar la validez y efectuar un testeo mínimo, la organización OCLC pone a disposición una componente de software en línea que permite testear en forma básica el comportamiento de la presente implementación [OCLC1]. Dicha comprobación ha resultado superada.

4.Servicios que es posible brindar con esta herramienta.

Si bien existe suficiente bibliografía comparativa entre la interoperabilidad de uno y otro protocolo [SYL], haremos una leve introducción: A diferencia del protocolo OAI, en el cual el cliente cosecha las referencias, el protocolo Z39.50 está orientado a la búsqueda y recuperación de la información referencial. Lo que diferencia uno y otro es el lugar donde se realiza la búsqueda y donde se almacena el repositorio de información.

El abanico de posibles servicios que es posible brindar con el uso de la presente herramienta, es realmente amplio: Una de las posibilidades y dado el dispar nivel de soporte informático con que las bibliotecas cuentan, consiste en ofertar publicar y difundir la información referencial, principalmente la relacionada a producción propia de cada facultad a través del protocolo SRU/W. La implementación de dicho servicio implica o bien recibir los registros bibliográficos de las mismas, o en su defecto acceder a los repositorios a través de alguna API o incluso al catálogo a través de web-scraping (fuerza bruta, solo en casos estrictamente imprescindibles). A los efectos del metabuscador cliente, el resultado siempre será XML (y preferentemente MARC-XML en el caso de información referencial bibliográfica).

Bajo es idea, la responsabilidad de SeDiCI no se limitaría a la Difusión de las contribuciones que los individuos o unidades académicas relacionadas con objetos digitales, sino a la información referencial (principalmente la de producción propia) que las bibliotecas poseen. Esa Difusión puede usarse con fines tales como la referencia o el descubrimiento de material relevante o de nuevo orientando hacia el sector de Procesos Técnicos en las bibliotecas, en el marco de un proyecto de catalogación por copia.

Sección 7.

Conclusiones.

Como conclusiones del presente trabajo podemos destacar las siguientes:

1. Es esperable que los sistemas de gestión de bibliotecas adopten una interfaz alterna de acceso a su información a través de protocolos "web friendly" como SRU/W. Esta tendencia debe ser aún mas fuerte en sistemas de gestión basados en herramientas OSS tales como PhpMyLibrary, EverGreen, Koha-BDF los cuales al no tener una arquitectura cerrada ni basada en productos propietarios facilitan la implantación de estos servicios. Esta adopción hará sensible a los mismos de resultar objeto de metabúsquedas en servicios internos o externos (Ej:Google).
2. Las plataformas de Bibliotecas Digitales, como es el caso de CELSIUS-DL deben contemplar la adhesión también a dichos protocolos. En el caso de las plataformas que resultan OAI Data Provider, dicha implementación es relativamente poco costosa de realizar, particularmente si el schema a devolver es Dublin Core. El nivel de complejidad crece en forma exponencial para la implementación completa de los servicios y combinaciones posibles contempladas en CQL, sin embargo posibilitan búsquedas en extremos pertinentes y precisas.
3. En un esquema de información distribuida, en el cual existen varios repositorios de información documental con escasa o nula intersección entre ellos (biblioteca, biblioteca digital, repositorios institucionales por ejemplo de pre-prints, información cosechada, etc) es esencial contemplar la máxima cantidad de protocolos de acceso, o al menos mínimamente un protocolo común neutro en términos tecnológicos, basado en XML. No basta con poner un enlace en el OPAC, indicando las bases de datos que se disponen o los "otros recursos". Cuantos e-usuarios se pierden por búsquedas fallidas de recursos bibliográficos, que se hallaban en otros repositorios locales?
4. Es esperable que así como Google adopta el protocolo OAI para Google Maps [Google] y de esa forma habilitar la indexación en dicho portal de recursos digitales, proyectos de similar orden, como Google Books adopten SRU/W para búsquedas o al menos para acceso a las publicaciones existentes en el marco académico. La iniciativa del proveedor de contenidos adoptando SRU para el acceso a sus publicaciones editadas es una señal sumamente positiva en dicha dirección [Jstor].

En el marco de las metabúsquedas en un servicio de referencia, es importante destacar nuevamente que la "bala de plata" contra los problemas de acceso a la información en los catálogos aun no existe. Persisten los problemas de falta de normalización en el acceso a las existencias que el uso del protocolo, no regula.

Como último agregado y simplemente como una visión (polémica) acerca de lo esperable para los diferentes sistemas informáticos relacionados con bibliotecas, resulta interesante plantear si es que en el futuro los proyectos relacionados con bibliotecas digitales deben o no ofertar una interfaz web visible de acceso propia a los recursos, favorecer su uso a los usuarios, realizar marketing como proyecto o simplemente funcionar como un conjunto de webservices que proveen la implementación para que los diferentes servicios internos y externos los consuman.

Bajo esta idea, un proyecto de biblioteca digital, al menos desde su implementación informática, consistiría en un conjunto de herramientas subyacentes que brindan entonces información a los OPAC's de las bibliotecas académicas cuando estas lo requieren, o pueden ofertarlo (sin pedidos de por medio) en forma de canales de noticias RSS. Conservarían información tales como las relacionadas con el copyright o derechos de acceso al documento en el acceso a los objetos digitales e implementarían todas las interfases de acceso (no destinadas al usuario final) que metabuscadores comerciales e institucionales puedan llegar a necesitar.

Siendo que como es natural, un proyecto de Biblioteca Digital involucra la tarea de varios especialistas en su formación y mantenimiento, para dar el caso en Bibliotecas Electrónicas de Tesis y Disertaciones [PAV], puede resultar dificultosa la obtención de recursos siendo que en realidad se apunta a que no es imprescindible que la Biblioteca Digital posea, por ejemplo, un portal web de consultas como tal. Quizás entonces, resulte dificultoso mostrar a las autoridades académicas uno de los "productos" del proyecto, siendo que en realidad el mismo está embebido en otros, aportándole contenidos.

Mas aún, bajo la visión que es recomendable que el usuario final acceda solo a un único punto de consulta, resultaría desaconsejable que los usuarios finales accedan a un portal propio de "Biblioteca Digital", siendo que sus servicios son "consumidos" por el catálogo de la biblioteca. Dejamos en este sentido, el debate abierto.

Sección 8.

Bibliografía y recursos.

[Moff] 'Marketing' with Metadata - How Metadata Can Increase Exposure and Visibility of Online Content. Version 1.0 8th March 2006 M.Moffat, PerX Project Officer. http://www.icbl.hw.ac.uk/perx/advocacy/exposingmetadata.htm#casestudy_jstor

[SYL] SRW/U with OAI.Expected and Unexpected Synergies. Sanderson, Young & Le Van. D-Lib Feb 2005. Vol 11 No. 2. <http://www.dlib.org/dlib/february05/sanderson/02sanderson.htm>

[OLD] Php2OAI. Universidad de Oldemburgo. <http://physnet.uni-oldenburg.de/oai/>

[RUS] Ruslan Z3950 gateway <http://www.ruslan.ru:8001/z3950/gateway.html>

[YAR]Yet Another Repository <http://yar.sourceforge.net/>

[PREBI] Proyecto de Enlace de Bibliotecas – UNLP <http://www.unlp.istec.org/prebi>

[CEL] Celsius Software. <http://celsius.prebi.unlp.edu.ar>

[ISTEC] Iniciativa Lib-Link <http://dll.istec.org>

[OCLC1] SRU Server Testing Page. <http://alcme.oclc.org/srw/SRUserverTester.html>

[OCLC2] OAI Harvester 2 Java Library <http://www.oclc.org/research/software/oai/harvester2.htm>

[LOGG] Library Of Congress Z39.50 gateway. <http://www.loc.gov/z3950/>

[CAN] Canadian GeoScience Knowledge Network. http://cgkn.net/cur/services/zserver_e.html

[FIS] Biblioteca del Departamento de Física de la UNLP <http://biblio.fisica.unlp.edu.ar>

[KOH] Koha-BDF. Branch del desarrollo original de Koha. <http://biblio.fisica.unlp.edu.ar:8080>

[ISM] IsisMarc. Plataforma de catalogación en formato MARC21 para tecnología Isis.

http://portal.unesco.org/ci/en/ev.php-URL_ID=11041&URL_DO=DO_TOPIC&URL_SECTION=201.html

[YAZPROXY] [SSE] [YAZ] <http://www.indexdata.dk>

[CAN] Canadian GeoScience Knowledge Network http://cgkn.net/cur/services/zserver_e.html

[SIU] Sistema de Información Universitario. Ministerio de Educación, Ciencia & Tecnología. República Argentina. Curso de Capacitación a distancia en el formato MARC21. <http://www.siu.edu.ar>

[Google] Google Maps y vía de indexación para OAI data Providers <https://www.google.com/webmasters/sitemaps/docs/en/other.html>

[ZIN] Página oficial de la iniciativa ZING. Z39-50 New generation. <Http://www.loc.gov/z3950/agency/zing/>

[SRU] Página oficial del protocolo SRU en Library Of Congress. <http://www.loc.gov/standards/sru/>

[SRW] Página oficial del protocolo SRW en Library Of Congress. <http://www.loc.gov/standards/sru/srw/>

[JSTOR] Jstor XML gateway http://www.jstor.org/about/xml_gateway.html

[PAV] ETD Guide. <http://www.etdguide.bibliored.cl/>