

Experiencias desde SeDiCI

Tecnologías para propagar los contenidos de una Biblioteca Digital

Ing. Marisa R. De Giusti
Directora PrEBi SeDiCI
Marmonti, Emiliano
Sobrado, Ariel
Vila, María Marta
Villarreal, Gonzalo Luján

- Iniciativa creada en la UNLP en el marco del Proyecto de Enlace de Bibliotecas (PrEBi).
- Objetivo prioritario: socializar el conocimiento generado en las diferentes areas académicas de la UNLP con el fin de devolver a la comunidad los esfuerzos destinados a la Universidad Pública

- Este objetivo engloba otros objetivos dentro, entre los que se destacan:
 - Incorporar a la UNLP al conjunto de Universidades que presentan su creación intelectual en forma abierta, presentando todo el material que se genera a la comunidad nacional e internacional y vinculando a sus creadores con docentes, estudiantes e investigadores en temas afines.
 - Fomentar una cultura de compartición de creaciones en un espacio común a todas las disciplinas
 - Incorporar a la UNLP a otras redes de recursos digitales existentes en la actualidad

¿En que consiste SeDiCI?

- Los autores incorporan sus creaciones al portal de SeDiCI, las cuales son catalogadas y puestas a disposición del público en general (previa autorización del autor)
- Se desarrolló un software llamado Celsius DL que permite por un lado la administración completa de SeDiCI (usuarios, materiales, catalogación) y por el otro la generación de las páginas del portal Web.

- La Iniciativa Archivos Abiertos (Open Archives Initiative - **OAI**) nace con el objetivo de desarrollar estándares de interoperabilidad para la difusión de contenidos en Internet
- Es un protocolo basado en XML
- Actualmente participan Universidades e Instituciones de todo el mundo
- Permite el acceso a un gran número de materiales digitales y publicaciones electrónicas: tesis en línea, informes, reportes, artículos, imágenes, etcetera.
- El protocolo define 2 roles:
 - Proveedor de Datos (Data Provider): mantiene uno o más repositorios que soportan el protocolo OAI-PMH como un método para exponer metadatos
 - Proveedor de Servicios (Service Provider) : realiza requerimientos a Data Providers

- Se implementó un software de recolección basado en la librería harvester OCLC.
- Los usuarios puede realizar búsquedas.
 - Sobre los documentos alojados en el servidor.
 - Sobre todo el material obtenido mediante el harvesting.
- La selección de los destinos es dinámica y se basa en el feedback con los usuarios.

- SeDiCI es Data Provider permitiendo el acceso a toda la información de los documentos alojados localmente.
- En una primer experiencia, los XMLs de los documentos eran generados estáticamente.

Ventajas: Acceso más rápido , sencillo , costos de implementación bajos.

- Luego se instaló la herramienta YAR la cual generaba una versión comprimida del XML obteniendo archivos de formato XMLType. Esta herramienta además provee una interfaz sencilla OAI-PMH .

Ventajas: Poco tiempo, costos bajos de implementación y semi-automatización de tareas.

- Debido al constante ingreso de documentos, era necesario volver a generar los XMLs y el archivo XMLType constantemente.
- El YAR no permitía la división por Sets (en SeDiCI cada uno representa un tipo de documento).

- Se busco una solución a este problema, implementando un nuevo Data Provider
 - Basado en una librería desarrollada por la Universidad de Oldemburgo en Alemania llamada **PhpOAI2**.
 - Genera la información dinámicamente.
 - Implementa en su totalidad el protocolo Open Archives admitiendo todas sus facilidades (incluyendo la división por Set)

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

- 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.
- Los autores de los trabajos entran en contacto con investigadores en temas afines en otras partes del mundo, se vinculan con otros círculos científicos de distinta complejidad y principalmente, se evita la duplicación de esfuerzos permitiendo mejorar el desarrollo de temas de investigación localmente.

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

- La UNLP posee actualmente 15 unidades académicas y en cada una de ellas se realiza un sin número de desarrollos heterogéneos.
- Se desarrollo de una plataforma ad-hoc (Celsius DL) capaz de soportar cualquier trabajo sin importar su formato o la disciplina a la que pertenece .Entre estos se tienen:
 - Publicación de revistas científicas .
 - Editorial de Libros Electrónicos. Generación y publicación de eBooks.
- A estos materiales se los puede acceder de forma ordenada a la información de cada uno, a las partes y a los documentos que lo componen.

- Permiten que dos equipos se comuniquen entre sí por medio de un protocolo llamado SOAP.
- Simple, eficiente y permite abstraerse de la plataforma e incluso del lenguaje de programación sobre el que se trabaje.
- Se utiliza un archivo de configuración WSDL, en formato XML, que indica explícitamente con que información se cuenta, que datos se presentan, de que tipo es cada uno, y que servicios que proveerán.
- Esta tecnología esta siendo muy utilizada en el área de bibliotecas digitales principalmente para la intercomunicación entre catálogos en línea, permitiendo así aumentar la oferta de información al usuario final.

- 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.
- Los visitantes de dichos catálogos pueden recibir ofertas de información tanto local (del catálogo) como información alojada en SeDiCI.

- Caso de estudio: Webservices para comunicar la Biblioteca de Física de la UNLP con SeDiCI.
 - El OPAC de la Biblioteca de Física de la UNLP (basado en el software Koha-BDF) se comunica con SeDiCI y obtiene de allí todos los documentos que pertenecen al esa área que están almacenados en SeDiCI.
 - Los usuarios de la Biblioteca también pueden realizar búsquedas sin la necesidad de salir de dicho portal ya que, por medio de los webservices.
 - 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.

Tecnologías para propagar los contenidos de una Biblioteca Digital

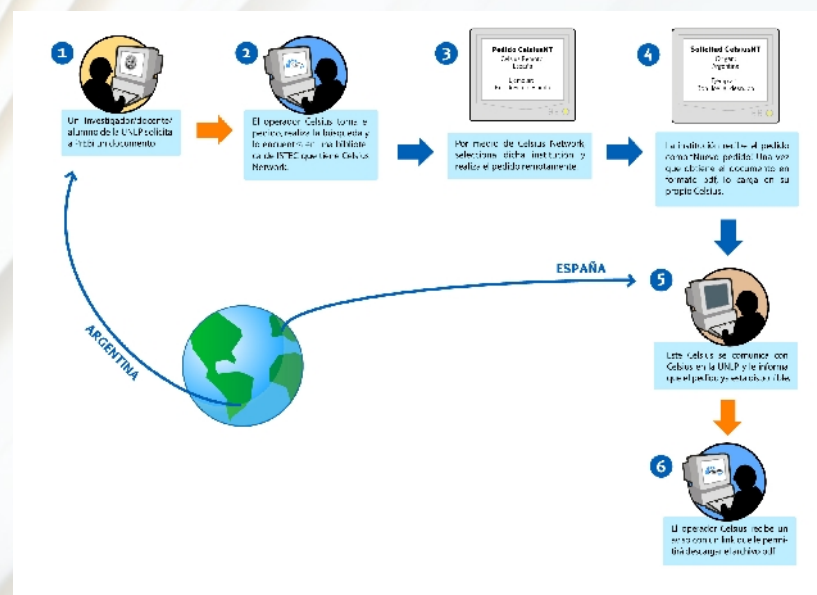
The screenshot shows a web browser window with the address bar displaying `http://biblio.fisica.unlp.edu.ar/cgi-bin/koha/opac-main.pl`. The page title is "Biblioteca del Departamento de Física - UNLP" and it includes contact information for the Faculty of Exact Sciences. The main content area features a search interface with a search bar, a dropdown menu for "Biblioteca" (set to "Todas"), and search options like "Cualquier palabra" and "Frase completa". There are also sections for "Novedades (Ultimos documentos)" and "Otras opciones" including "Bibliografía por Carrera o Curso" and "Historial de Búsquedas". On the right side, there is a login section titled "Ingreso de Usuarios" with fields for "Usuario" and "Password", and a list of "Canales de noticias" (Física, Matemática, Química) and "Recursos en línea" (22 Tesis de Postgrado, 1 Disertación, etc.). A sidebar on the left contains a navigation menu with items like "La Biblioteca...", "Infraestructura", "Catálogos", and "Reglamentaciones".

- Otro uso de webservices en software para Bibliotecas.

–La plataforma **Celsius** para gestión de pedidos del Proyecto de Enlace de Bibliotecas (PrEBi) esta evolucionando hacia la tecnología de webservices.

–Las distintas instancias de Celsius instaladas se comunicaran entre sí, automatizando muchas tareas donde era necesaria la intervención directa de los operadores.

–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.



- Estándar internacional para la comunicación entre sistemas informáticos.
- Cada vez más importante en el desarrollo de sistemas en el ámbito de las Bibliotecas Digitales.
- Provee un método eficiente para recuperar información basado en la filosofía Cliente-Servidor.
- Permite abstraerse del hardware y del software.
- 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.

- Problemas:
 - la norma puede ser algo compleja.
 - Requiere la apertura de puertos TCP/UDP. No todos normalizan al puerto 210, lo que implica convencer a los webmasters acerca de las necesidades técnicas de los puestos de catalogación.
- YazProxy.
 - herramienta desarrollada por Indexdata.
 - Soluciona los problemas de apertura de puertos desde las múltiples computadoras destinadas a procesos de catalogación.
 - El transporte se realiza sobre http.
 - Utiliza XML como envoltorio.

- ZING: Z39.50 International Next Generation
- Es un proyecto de la Z39.50 International Maintenance Agency
- Promueven un conjunto de iniciativas con el objetivo de hacer más accesibles los contenidos tanto intelectuales cómo semánticos del protocolo para transformarlo en una herramienta mas atractiva para desarrolladores, usuarios, proveedores de información y distribuidores

- CQL: Common Query Language
- Lenguaje formar para representar consultas
- Premisa: las consultas deben ser:
 - Fáciles de leer y de escribir
 - Intuitivas
 - Mantener la expresividad de otros lenguajes más complejos (SQL, XQuery, PQL)

- Algunos ejemplos simples de CQL:
 - title = "complete dinosaur."
 - El título contiene.
 - title exact "the complete dinosaur."
 - El título ES exactamente.
 - dinosaur and "ice age."
 - Registros en los que coinciden ambos criterios.
 - publicationYear < 1980.
 - Publicado antes de...

- Ejemplos más complejos:
 - (Caudal or dorsal) prox vertebra : *registros en los cuales las palabras caudal o dorsal estan cercanos a la palabra vertebra.*
 - title any "dinosaur bird reptile" : *el titulo contiene cualquiera de las palabras dinosaur, bird o reptile.*
- Más ejemplos en <http://zing.z3950.org/cql/index.html>.

- SRU: Search / Retrieve Via URL.
- Protocolo estándar de búsqueda sobre Internet.
- Las consultas se expresan en CQL.
- Principal ventaja: fácil y sencillo.
- Un servidor recibe un requerimiento SRU, verifica y genera una respuesta en formato XML.
- Un requerimiento SRU es una url http que consta :
 - url base: dirección y puerto del servidor.
 - parámetros: clave=valor separados por &.
 - Ejemplo:
[http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur.](http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=dinosaur)

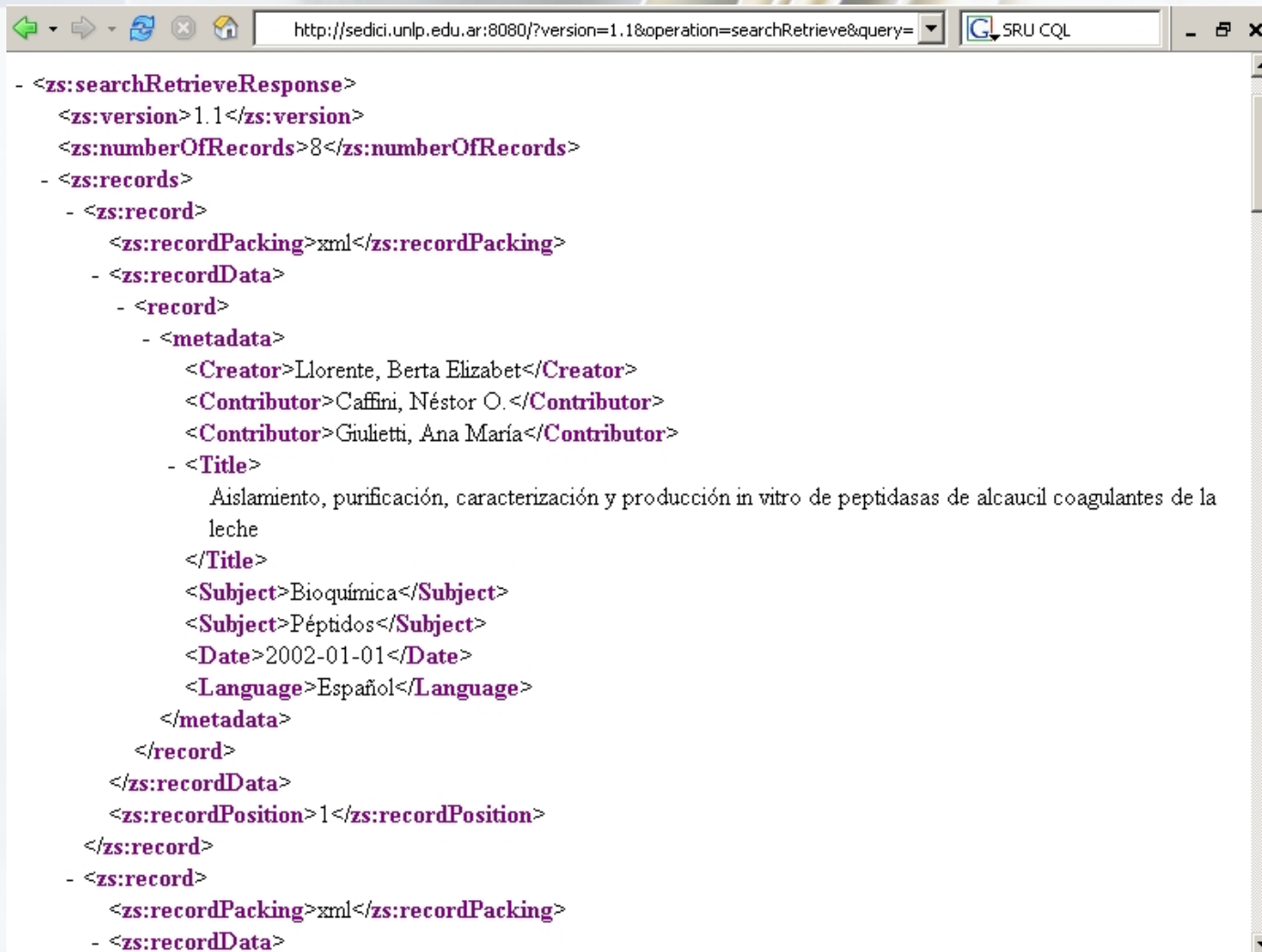
- SRW: Search/Retrieve Web Service.
- Es una variación de SRU.
- Los mensajes son enviados en el formato XML sobre el protocolo HTTP , a través de SOAP.
- Esta diferencia muy significativa: mejor soporte para extensión , autenticación y características de webservices.

- Los protocolos SRU/SRW proveen mayor claridad en la implementación, lo que implica mayor extensionalidad
- Están basados en un estándar, lo que aporta
 - Funcionalidad verificada
 - Documentación

- Se plantearon los siguientes objetivos:
 - Analizar si dicho protocolo es posible de implementar con las herramientas tecnológicas OSS disponibles.
 - Analizar los costos necesarios para implementar y el costo en esfuerzo de adaptación al repositorio local de metadatos administrado por SeDiCI.
- El repositorio de metadatos de SeDiCI esta basado en un motor de base de datos relacional Open Source.

- Se instaló el producto **Index Data simple Server**, el cual implementa el acceso a través del protocolo Z y de los protocolos SRU y SRW por medio de la librería yaz.
- Se implementaron los servicios básicos de acceso y consulta simples.
- Se provee el acceso a la información referencial existente en el repositorio principal de **producción propia**.
- Celsius-DL admite mapeos hacia recordSchemas tales como MARC-XML, MODS y Dublin Core.
- SeDiCI pasa a ser un servidor Z39.50 y un “implementador” del servicio SRU/SRW.

- ¿Cómo acceder al repositorio de SeDiCI via SRU?
- La URL correspondiente es: <http://sedici.unlp.edu.ar:8080?version=1.1&operation=searchRetrieve&query=Bioquimica&maximumRecords=&recordSchema=dc>.
 - version = versión del protocolo.
 - operation= tipo de operación contemplada en SRU/W.
 - query=expresión de consulta (todos los resultados que tienen Bioquimica en alguno de sus metadatos).
 - maximumRecords=<x> cantidad de registros a visualizar por búsqueda.
 - recordSchema=se refiere al esquema XML a devolver. En el caso de SeDiCI solo DC.



The screenshot shows a web browser window with the address bar containing the URL: `http://sedici.unlp.edu.ar:8080/?version=1.1&operation=searchRetrieve&query=`. The browser title is "SRU CQL". The main content area displays an XML response from a search service, with the following structure:

```
- <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>
```

- Con OAI, el cliente cosecha las referencias y las almacena localmente.
- Con Z39.50 el cliente solicita la búsqueda y recuperación de la información referencial a un servidor remoto.
- 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.

Muchas gracias por su atención.

<http://sedici.unlp.edu.ar>

info@sedici.unlp.edu.ar

- Agradecimientos especiales:
 - A Gisele Jaquenod por el diseño visual y por la dedicación en éste y todos los trabajos.
 - A Emiliano Marmonti por su constante cooperación
 - A Ariel Lira por sus aportes, críticas e ideas