Prepared for delivery at the 2004 meeting of the Latin American Studies Association, Las Vegas, Nevada, October 7-9, 2004

Bibliotecas Digitales Latinoamericanas en el Marco de OAI-PMH

María Inés Bravo, Facultad Latinoamericana de Ciencias Sociales (FLACSO), Chile, biblioteca@flacso.cl

Kent Norsworthy, Latin American Network Information Center (LANIC), University of Texas, Austin, kentn@mail.utexas.edu

Paula Pardo Lorca, Facultad Latinoamericana de Ciencias Sociales (FLACSO), Chile, ppardo@flacso.cl

1. Presentación

El **Portal de Archivos Abiertos de América Latina** es un proyecto que se inserta dentro del Proyecto de Recursos de Investigación sobre Latinoamérica - LARRP, en colaboración con el Latin American Network Information Center - LANIC.

El objetivo principal del proyecto es proporcionar metodologías y montar una infraestructura que permitirá a instituciones latinoamericanas, que producen información para la investigación en las ciencias sociales, difundir más ampliamente sus contenidos digitales por Internet. Con esto, se pretende proporcionar una mejor base de información a los investigadores, tanto dentro como fuera de la región, en lo que se refiere a la clase de publicaciones de escasa circulación conocida como *Literatura Gris*.

Esto se propone lograr en el marco de la **Iniciativa de Archivos Abiertos** (Open Archives Initiative, OAI). La OAI es un modelo descentralizado que promueve la interoperabilidad o compatibilidad entre grupos de bibliotecas digitales unidos en torno a una comunidad de interés, en este caso la de Estudios Sobre América Latina. Estas bibliotecas, conocidas en la jerga de la OAI como *Proveedores de Datos*, exponen metadatos de sus publicaciones a un cosechador o *Proveedor de Servicios*, que se encarga de proporcionar a los usuarios servicios avanzados de búsqueda que permitan indagar simultáneamente los contenidos en texto completo del conjunto de bibliotecas digitales participantes.

Durante la actual etapa, el proyecto esta trabajando en la elaboración de metodologías y procedimientos para el escaneo y digitalización de contenidos, para la generación y conversión de metadatos, y para el montaje del cosechador. Trabajando estrechamente con personal de FLACSO-Chile, hemos concluido una primera etapa con el montaje de un proveedor de datos en prototipo, que incluye metadatos convertidos del catálogo existente de FLACSO, que son extraídos por medio de un Cosechador y un Static Repository Gateway (ver abajo) instalados en LANIC. En la siguiente etapa, pretendemos aumentar el número de publicaciones de FLACSO disponibles, pasar el prototipo a un ambiente de producción, y replicar la experiencia con otras instituciones de la región.

En esta ponencia, pretendemos dar una visión introductoria de la OAI, los antecedentes de nuestro proyecto, y una descripción de sus principales componentes y logros hasta la fecha.

2. Antecedentes Regionales

Este proyecto está concebido en base al análisis de un conjunto de factores propios del mundo de la producción y circulación del conocimiento en el ámbito académico (conocido en inglés como la “scholarly communication”). Algunos de estos factores giran en torno a la llamada “comunicación Sur-Sur”, en este caso, el flujo de ideas y propuestas entre investigadores de los diversos países de América Latina. Otros están relacionados con la “comunicación Norte-Sur”, sobre todo en lo que se refiere a la creciente demanda en los países del Norte por la información producida en, y sobre América Latina.

Específicamente, existen dos tendencias que se han dado en los últimos años que proyectos como este están diseñados a enfrentar. Por un lado, ha habido una tremenda expansión en la cantidad de información y análisis producida y “publicada”, sobre todo electrónicamente, en el seno de las ciencias sociales latinoamericanas. En muchos casos, se trata de valiosos avances de investigación u otro tipo de publicación que en la era pre-digital, no hubiera tenido posibilidad de llegar a ser “publicado” por el elevado costo de producción “en papel”.

Por otro lado, también ha habido una demanda creciente dentro de la región por la información publicada por otros investigadores latinoamericanos. Esta tendencia nace producto de varios factores, entre ellos un proceso de relativa “maduración” en varios campos de estudio que trae mayores niveles de teorización autóctona dentro de la región, y la creciente tendencia de realizar estudios comparativos entre varias naciones de la región.

Estas dos tendencias se han dado a la par de la aparición de Internet y la Web dentro de América Latina, no solo en el ámbito comercial, sino también en el sector académico y como herramienta de comunicación y búsqueda de información. Es en este sentido que se ha visto el fenómeno de un número cada vez mayor de centros de investigación, algunos independientes, otros afiliados a las universidades, que cuentan con su propia presencia en Web y que “publican” sus avances de investigación y otros documentos en su sitio.

El problema para el que busca la información es que estos documentos generalmente resultan sumamente difíciles de recuperar. Tradicionalmente, el usuario tenía dos alternativas para recuperar este tipo de información. Una era por medio de un buscador global, como Google o AltaVista. Aparte del problema de que muchas de estas publicaciones permanecen “ocultas” ante estos buscadores, de cualquier manera la búsqueda de este tipo de información especializada muchas veces resulta poco eficiente desde el punto de vista del tiempo que se requiere, así como de la calidad desigual de recursos que pueden ser recuperados. La segunda alternativa era buscar las publicaciones “centro por centro,” es decir, visitar los sitios Web de cada centro que pudiera tener

publicaciones de interés. Al igual que con el buscador global, esta modalidad puede resultar muy frustrante y complicada.

Es en este contexto que la OAI ofrece una metodología que facilita a los centros la “auto-publicación” de sus materiales intelectuales, y a los usuarios la recuperación de las mismas por medio de buscadores basados en los metadatos, que son de mucho más detalle y sofisticación que los buscadores tradicionales.

3. Antecedentes Extra-Regionales

Como región, América Latina ha sido objeto de investigación y estudio con cada vez más fuerza en los últimos años. Este creciente interés no es solamente el producto de fuerzas coyunturales como la globalización, el libre comercio o la integración económica, sino que nace también de profundas raíces y enlaces históricos y culturales. A su vez, la creciente facilidad de la comunicación y el transporte ha contribuido a masivos flujos migratorios, entre ellos el que nos ha llevado a una situación en que hoy la población de origen latino en Estados Unidos está cerca de convertirse en la principal “minoría” dentro del país. Las conexiones hemisféricas que nos unen son de carácter político, pero también son intelectual y personal; son de carácter comercial, pero también son cultural. El resultado es tan obvio e importante en el ámbito académico como en cualquier otro, evidenciado por el cada vez mayor número de estudiantes de español en las aulas universitarias, por el crecimiento de los programas de estudio sobre América Latina, y por las actividades cada vez mas importantes de las organizaciones profesionales del campo, como es el Latin American Studies Association (LASA).

El campo de los Estudios Sobre América Latina incorpora dimensiones académicas, estratégicas, didácticas, comerciales, culturales y cívicas. Cada una de estas dimensiones requiere de una base sólida de información, base que las grandes bibliotecas de investigación de Estados Unidos han estado construyendo desde hace por lo menos un siglo. A estas alturas, las colecciones tradicionales de libros son bastante fuertes. La producción y la demanda por información y análisis publicado en otros formatos también ha crecido enormemente, pero en este caso, la respuesta de las bibliotecas de investigación no ha sido tan impresionante. Cada vez más, los estudiantes e investigadores buscan recursos de información en formatos no-tradicionales al tiempo que diversas organizaciones latinoamericanas, operando al margen de los circuitos comerciales establecidos, han aumentado significativamente su oferta de materiales. Nuevos formatos de difusión, sobretodo las publicaciones electrónicas, también han hecho más difícil las tareas de trabajar con las fuentes de información, ya sea desde el punto de vista de su identificación, la provisión de acceso a ellas, y su preservación.

Desde ya, la comunidad de bibliotecarios latinoamericanistas se ha movilizado para hacer frente a estos retos. Una de las principales respuestas ha sido el lanzamiento de la Latin Americanist Research Resources Project (LARRP), un consorcio iniciado en 1994 bajo el patrocinio conjunto de la AAU y ARL. Actualmente, LARRP cuenta con 48 miembros institucionales en Estados Unidos y Canadá, así como con siete participantes latinoamericanos. Las primeras iniciativas de LARRP, realizadas dentro de Estados Unidos, incluían la digitalización de mas de 75 mil páginas de discursos presidenciales de Argentina y México; la creación de una base de datos de tablas de contenido de revistas especializadas latinoamericanas (LAPTOC), que cuenta actualmente con registros para mas de un cuarto de millón de artículos; y la confección de mecanismos voluntarios para que las bibliotecas participantes pudieran reacomodar sus presupuestos para la adquisición de monografías por medio de un esfuerzo coordinado, mientras simultáneamente se fortalecería la base nacional de títulos en existencia.

Con financiamiento del gobierno federal de Estados Unidos, bajo el Programa de Innovación Tecnológica y Cooperación para Acceso a Información Extranjera, (TICFIA), ciclo 1, hemos hecho grandes esfuerzos por modernizar y expandir a LAPTOC. Estos incluyen el rediseño total de la interface para el usuario, facilitando y agilizando las consultas a la base de datos con pasos más sencillos y funcionalidades adicionales. Estos cambios han sido implementados cuidadosamente, para asegurar cumplir con los nuevos requerimientos técnicos para las transacciones del sistema de Préstamo Inter bibliotecario (ILL). Al mismo tiempo, hemos expandido la cobertura de la base de datos de los primeros tres países en el período piloto, a la casi totalidad de los países de la región por medio de convenios con varias instituciones latinoamericanas. El trabajo colaborativo con nuestros socios latinoamericanos ha sido vital para comprobar la eficacia y los beneficios de la cooperación internacional para facilitar el acceso a la información producida en América Latina, elemento clave en la concepción del segundo ciclo de TICFIA, como veremos más abajo.

A pesar de estos y otros logros, sigue una laguna en lo que se refiere al acceso a un tipo de publicaciones: la llamada *literatura gris*, es decir, los documentos de trabajo y avances de investigación publicados por ONGs y centros de investigación. Al mismo tiempo, estamos luchando para encontrar sistemas eficaces para sistemáticamente identificar, capturar, y preservar la cada vez más voluminosa producción de información electrónica confeccionada dentro de la región. El proyecto LAOAP, con el patrocinio de una segunda ronda de financiamiento TICFIA, o sea TICFIA2, pretende enfrentar estos dos retos por medio del establecimiento de nuevas alianzas con instituciones latinoamericanas y de una expansión de la infraestructura técnica implementada en el ciclo TICFIA1.

El objetivo principal del proyecto es facilitar el acceso a la *literatura gris* latinoamericana, o sea todos aquellos materiales que son publicados en papel o

electrónicamente por agencias o instituciones dentro de la región, pero que no ingresan al circuito de comercialización tradicional y por tanto, son difíciles de conseguir para las bibliotecas y para los investigadores tanto dentro, como fuera, de la región. También busca enfrentar el problema, cada vez más sentido en Internet, de que muchas de las publicaciones electrónicas que han sido difundidas por medio de la Web aún son muy difíciles si no imposibles de encontrar, por la falta de sistemas adecuados de descripción, acceso, búsqueda y preservación. Otra característica de esta literatura es que suele promover puntos de vista alternativos o no tradicionales en lo político y en lo intelectual. También, tiende a presentar las teorías y los hallazgos de investigación más de punta.

En la mayoría de los casos, cuando se trata de la literatura gris, los derechos de autor en torno a la distribución o la digitalización no supone un problema, pues generalmente con este tipo de publicaciones los autores están comprometidos con compartir los hallazgos de sus investigaciones o de divulgar lo más ampliamente posible sus puntos de vista políticos.

Algunos ejemplos de lo que se incluye cuando hablamos de la literatura gris serían: reportes (por ejemplo, borradores o versiones preliminares y avances de investigación, reportes técnicos, reportes estadísticos, etc.); tesis de postgrado; ponencias de conferencias y otros foros; especificaciones y estándares técnicos; traducciones no-comerciales; bibliografías; documentación técnica y comercial; y documentos oficiales que no son publicados comercialmente (por ejemplo, informes y otros documentos primarios de agencias gubernamentales y paraestatales).

El proyecto LAOAP, también conocido como TICFIA2, que está siendo ejecutado en un período de tres años, tiene dos componentes básicos:

1) ***Digitalización de Publicaciones y conversión/creación de Metadatos***. El objetivo es crear una red de bibliotecas digitales (*Proveedores de Datos* en el lenguaje OAI). Las instituciones que desean integrarse a la red tendrían que contar con publicaciones electrónicas (por ejemplo, en formato PDF) o en su defecto escanearlos, y metadatos sobre dichas publicaciones.
2) Montaje de un ***Portal Latinoamericano de Recursos para la Investigación*** y servicios avanzados de búsqueda en LANIC. Por medio de este Portal, los usuarios podrán efectuar búsquedas simultáneamente en todos los registros de metadatos del conjunto de bibliotecas digitales participantes en el proyecto.

Ampliándose sobre estos dos puntos, la modalidad de trabajo que se adoptó para el primer acápite, la digitalización y los metadatos, ha sido la de “efecto de demostración”. Específicamente acordamos trabajar, en primera instancia, con dos instituciones de la región—la Facultad Latinoamericana de Ciencias

Sociales (FLACSO-Chile) y el Centro de Investigaciones Regionales de Mesoamérica (CIRMA-Guatemala)—para elaborar metodologías y escoger "mejores prácticas" en torno a la digitalización y la generación de los metadatos.

Hasta la fecha, personal de LANIC y de FLACSO-Chile han trabajado juntos en una serie de iniciativas en ese sentido, entre otros, la confección de un inventario detallado de publicaciones a ser tratados en el marco del proyecto; la evaluación y adquisición de equipos para el escaneo; pruebas de escaneo con documentos "prototípicos"; y la elaboración de procedimientos para la conversión de metadatos del catálogo online de FLACSO al formato Dublin Core, un estándar internacional de metainformación general, en un archivo XML (ver detalles en las siguientes dos secciones).

El montaje del Portal Latinoamericano contemplado en el segundo acápite es un proceso más prolongado. Sin embargo, paralelo al trabajo con FLACSO-Chile, personal de LANIC ha venido sistemáticamente montando la infraestructura necesaria para dar los servicios requeridos bajo el modelo OAI. El trabajo se puede resumir en la investigación, instalación, configuración, programación, y pruebas con prototipo con el Cosechador de Metadatos, respaldado por una base de datos para alojar los Metadatos y el *Static Repository Gateway* (SRG).

Los avances en estos dos acápites culminaron en marzo del 2004 con las primeras pruebas "en vivo" de recuperación de metadatos, que habían sido convertidos mediante el procedimiento automatizado generado por FLACSO, que están almacenados en un Static Repository y que fueron "cosechados" por medio del SRG de LANIC. El prototipo está disponible en el sitio Web del proyecto, en http://lanic.utexas.edu/project/laoap/.

El buscador del Portal se encuentra en desarrollo, pero ya se cuenta con un diseño también de prototipo para la interface de Búsqueda Básica, Búsqueda Avanzada, y Listar. Con cualesquiera de las tres variantes, el usuario debe primero escoger operar sobre uno, varias, o todas las colecciones o bibliotecas digitales agrupadas en el Portal. Con Búsqueda Básica, el usuario ingresa sus términos de búsqueda, y estos son rastreados en todos los campos de metadatos. La opción de Búsqueda Avanzada proporciona más flexibilidad, al permitir la búsqueda por campos específicos de metadatos. La tercera alternativa, Listar, permite al usuario recuperar una lista completa de todas las publicaciones de la colección escogida, con la opción de ordenar los resultados por autor, por título, o por tema.

4. FLACSO-Chile y el Papel de la Biblioteca

La Facultad Latinoamericana de Ciencias Sociales FLACSO, es un organismo internacional de carácter académico, regional y autónomo, creado en 1957 por los gobiernos de América Latina y el Caribe. Santiago de Chile fue la primera Sede Académica de FLACSO creada el mismo año de su fundación. Su objetivo es promover la enseñanza, la investigación y la cooperación científica en el campo de las ciencias sociales en toda América Latina y el Caribe. FLACSO constituye la red más importante de la región en su campo de conocimientos. Cuenta con diez Sedes y Programas en América Latina y el Caribe: México, Guatemala, El Salvador, Costa Rica, Cuba, República Dominicana, Ecuador, Brasil, Argentina y Chile.

Junto con la creación de FLACSO, se inician las actividades de la biblioteca institucional, teniendo a la fecha 47 años de existencia. A lo largo de su trayectoria, FLACSO-Chile se ha constituido en uno de los centros de mayor relevancia en el campo de la investigación y docencia de las ciencias sociales tanto en el país como en el resto de América Latina y el Caribe. Su continuidad de funcionamiento en el país radica en haber sabido combinar en las diversas circunstancias, excelencia académica, compromiso intelectual y una estructura organizativa estable, como a la vez flexible, que ha asegurado una gestión eficiente en su quehacer.

La misión de FLACSO-Chile se realiza a través de tres objetivos centrales de su trabajo académico:

* Contribuir, en el más alto nivel académico, al desarrollo democrático del país y la región, incorporando dimensiones internacionales al debate nacional.
* Difundir los resultados de estos procesos, estudios, análisis e investigaciones entre actores claves del desarrollo nacional y regional.
* Establecer programas de formación dirigidos a las nuevas generaciones de cientistas sociales.

Como objetivo general y misión, la Biblioteca de FLACSO-Chile busca satisfacer las necesidades de información bibliográfica de los académicos en las distintas áreas de estudio de la Institución, así como las de los organismos vinculados a las ciencias sociales del país y de la región, contribuyendo de esta forma a que la información disponible se transforme en conocimiento y acción. Orienta sus actividades estratégicas hacia el objetivo de brindar y proveer de información oportuna y actualizada a profesores y ayudantes de FLACSO, además de satisfacer necesidades de información de investigadores y estudiantes de pre y posgrado, a tesistas a nivel nacional e internacional interesados en conocer la realidad social y política de nuestro país y a otras organizaciones vinculadas a las áreas temáticas de la FLACSO, como también a entidades de gobierno, y a políticos y parlamentarios.

Una preocupación permanente de la biblioteca es dar a conocer a la comunidad académica y a la opinión pública las investigaciones y estudios publicados por la Institución y su esfuerzo en relación al desarrollo de las ciencias sociales en Chile y la región.

La Biblioteca cumpliendo su rol de gestora natural del conocimiento al interior de la Institución, y conciente del impacto que tiene la información validada, actualizada y oportuna, ha presentado diversos proyectos que le permitirán hacer frente a los desafíos del mundo moderno y sus exigencias en materia de acceso a la producción del conocimiento.

Es así como gracias al financiamiento de la Fundación Ford se pudo implementar el catálogo de la biblioteca en línea y además esto permitió la creación y puesta en línea en junio de 2003 del Fondo Documental Eugenio Ruiz-Tagle, dedicado al acopio, conservación y divulgación de documentos producidos por partidos políticos y organizaciones sociales chilenos entre 1964 y 1989, cumpliendo así con una de las estrategias de FLACSO en la promoción del conocimiento generado en la Institución.

En esta época de transición y cambios tecnológicos acelerados, nos enfrentamos a la necesidad de reestructurar y modificar nuestras funciones y servicios. De ahí la importancia que resulta para la biblioteca el adecuar y situar nuestro sistema bibliográfico dentro de las nuevas formas de transmisión del conocimiento.

En este marco de creciente aumento de la información y de interés por difundir la producción histórica de FLACSO surge otro proyecto. Es el del *Portal de Archivos Abiertos de América Latina,* también conocido como el proyecto TICFIA2 bajo el patrocinio de LARRP. Básicamente este proyecto permitirá digitalizar toda la colección de generación propia, es decir, documentos y libros, para ponerla a disposición de la comunidad académica internacional a través de un portal Web de acceso gratuito.

Este proyecto involucra la digitalización de todos los documentos de FLACSO-Chile desde el año 1975 en adelante, lo que cubre un total de:

* 824 Documentos de Trabajo
* 12 vols. de la revista *Cono Sur*
* 17 vols. de la revista *Fuerzas Armadas y Sociedad*
* 118 títulos de libros.

5. Procedimientos de homologación y conversión de campos de Micro ISIS a Dublin Core

El catálogo bibliográfico de FLACSO-Chile está soportado en la base de datos micro CDS/ISIS y estructurado de acuerdo al formato CEPAL. De acuerdo a los propósitos del proyecto fue necesario seleccionar los registros pertenecientes sólo a publicaciones FLACSO o aquellas coautorías que contaran con autorización de divulgación. Como resultado de esto, se seleccionaron 2.976 registros, analíticos y monográficos, los que formarán parte del portal y de los cuales, se está procediendo a recuperar los textos completos en formato digital. Para efectos de evitar confusiones en esta primera etapa del proyecto, se optó por separar los registros que irán al cosechador de la base de datos bibliográfica general, creando así una base que en estructura es igual, pero que contiene sólo los registros seleccionados para el portal.

Por otra parte, el cosechador de metadatos del proyecto trabaja sobre el estándar de Dublin Core, del cual, fueron seleccionados 12 de los 15 campos que lo componen. Para poder hacer compatible los registros del catálogo con las exigencias del formato fue necesario identificar los campos existentes en Micro ISIS y homologarlos con los del Dublin Core. La mayor dificultad de este paso se dio en casos en que los campos que para Dublin Core representan un solo campo, en Micro ISIS se encontraban dispersos en varios campos, subcampos u ocurrencias. Una vez homologados los campos, se procedió a crear un formato de salida de datos, en los cuales, cada uno de los campos quedaron etiquetados de acuerdo a Dublin Core (ver imágenes 1 y 2).

Afortunadamente, Micro ISIS permite la creación de un archivo de salida en texto plano (txt), formato de texto requerido para generar el archivo que posteriormente será accesado por el SRG y recuperado por el cosechador. En el proceso de generación del formato, una de las dificultades fue otorgar la fecha actualizada de exportación de los registros, debido al estándar que exige Dublin Core de presentar los datos en su forma año/mes/día. Esto se solucionó pensando en insertar textos “literales” en el momento de creación del archivo de salida. Otra dificultad fue otorgar un número unívoco a cada registro y que será el número de identificación que use el cosechador, en este caso, se optó por usar el número MFN que otorga Micro ISIS.

Imagen 1: Registro con formato de salida FLACSO

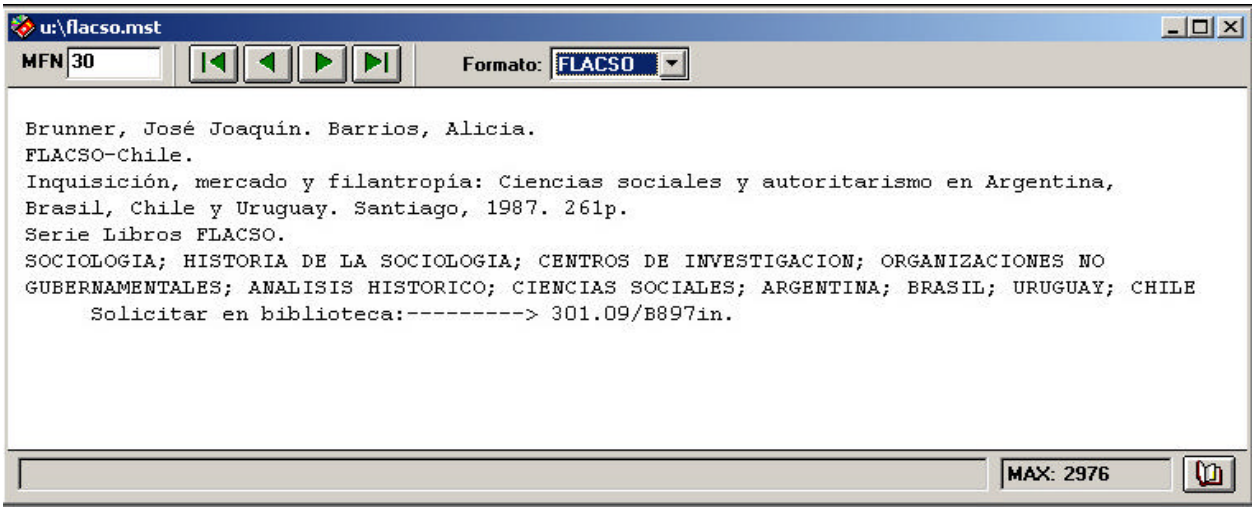
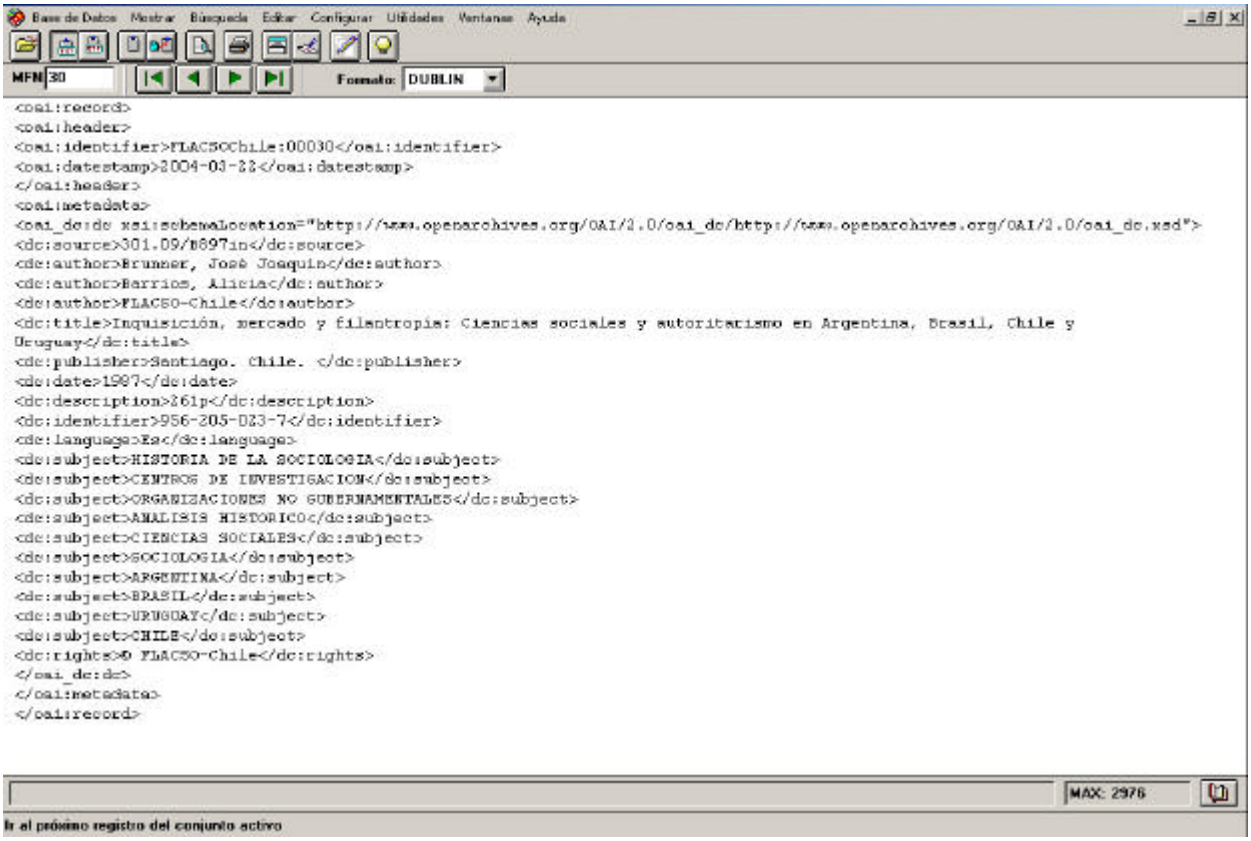


Imagen 2: Registro con formato de salida etiquetado de acuerdo a Dublin Core



Cuando este proceso estuvo listo, se generó un archivo de prueba con todos los campos de unos 30 registros, que fue enviado a LANIC para ser usado en el prototipo. Esta primera prueba fue bastante exitosa.

Respecto del alojamiento de los archivos con los textos completos, se probó, en primera instancia alojarlos en LANIC, para posteriormente subirlos a un servidor de FLACSO. Por lo tanto, actualmente el prototipo arroja los resultados con los registros que recupera el cosechador en LANIC y entrega un enlace al documento completo que se encuentra alojado en FLACSO.

6. Digitalización de la colección

En forma paralela al proceso de homologación de los registros, se puso en marcha el proceso de preparación de la colección para digitalizarla y transformarla a archivos PDF, el que se ha dividido en las siguientes actividades

* Selección de los documentos de acuerdo a los registros en la base de datos
* Preparación para escaneo
* Escaneo de los documentos
* Preparación del archivo final PDF
* Vinculación del archivo al registro correspondiente en la base de datos
* Subir archivos al servidor
* Enviar el archivo txt a LANIC para disponerlo para el SRG

En relación a la etapa de selección, actualmente, la biblioteca ha verificado y recuperado los impresos de la mayor parte de la colección (que consta de 2.976 registros a la fecha). Esta etapa del proceso ha requerido verificar registro por registro las existencias. Para los documentos, que por alguna razón no se encuentran físicamente en la institución (pérdida, destrucción por el tiempo, etc.) se le ha hecho un seguimiento y se ha pedido la colaboración de los autores para obtener una copia del original. Otro detalle que se ha cuidado, es obtener todos los permisos y autorizaciones de aquellos documentos de los cuales FLACSO no tiene todos los derechos, en este caso, nos referimos a las coautorías con otras instituciones. Cabe destacar al respecto, que la colaboración ha sido muy satisfactoria y que los autores se han mostrado muy receptivos con el proyecto.

La actividad de preparación de los documentos está bastante avanzada y ya se tiene un segmento de la colección lista para comenzar la digitalización. De acuerdo a los cálculos de Biblioteca, a partir de mediados de agosto de 2004 ya estaría instalada toda la infraestructura tecnológica que permitirá escanear la totalidad de la colección. Para este proceso se cuenta con un escáner profesional de última generación que permite digitalizar hasta 50 páginas por minuto. Este procedimiento de escaneo rápido se hará con todos aquellos

documentos que tengan duplicados y que por lo tanto, pueden ser guillotinados, esto para dar fluidez al proceso. Para aquellos documentos que no tengan respaldo impreso y se encuentren deteriorados, el proceso se hará manualmente en otro escáner de cama plana.

Para la transformación de los archivos a PDF se optó por utilizar las herramientas Adobe Acrobat 6.0, PDF Tools y Omnipage 10. Un punto importante, es que se está viendo la posibilidad de generar estos documentos mediante el método imagen+texto, que implica generar un archivo PDF que contiene la imagen del documento, pero a la vez recupera el texto (caracteres) del documento. Al tener los documentos creados en este formato, será posible recuperar en texto libre los contenidos de los archivos.

Se espera que cuando ya se tenga un porcentaje de la colección digitalizada, alrededor de un 10%, el equipo servidor de FLACSO que contendrá los archivos ya estará habilitado para proveer la conexión a los documentos completos. A partir de ese momento, se podrá enviar a LANIC una primera remesa de documentos para subir al prototipo y que podrán ser consultados y bajados cuando el portal comience su marcha blanca.

7. Cómo integrarse al proyecto

La pretensión del proyecto es que la experiencia arriba descrita para la creación del *data repository* por parte de FLACSO pueda ser replicado por otras instituciones que tienen interés en participar. Para este efecto, hemos confeccionado un esquema con varias alternativas de incorporación, descrito en esta sección, así como un breve cuestionario, que se puede consular en el Anexo “Formulario para Participantes”.

El esquema de la Iniciativa de Archivos Abiertos (Open Archives Initiative – OAI, por sus siglas en Inglés) tiene dos componentes básicos: un conjunto de *Proveedores de Datos*, que son las bibliotecas digitales ("repositorios" de objetos digitales y sus metadatos), y un *Proveedor de Servicios*, que incluye un "cosechador de metadatos" y ofrece un portal con servicios de búsqueda avanzada para usuarios.

Hay tres pasos o requisitos para aquellas instituciones que tienen interés en participar en el proyecto LAOAP como repositorio o biblioteca digital:

1. Contar con un acervo de documentos electrónicos (material digitalizado, como: textos, fotografías, grabaciones sonoras, video, etc.), o bien la capacidad de convertir los contenidos no-digitales, como por ejemplo por medio de operaciones de escaneo. Para textos, basta con poder generar archivos en formato PDF.

2. Acceso a un servidor Web propio, o en su defecto a un servicio permanente de alojamiento en Web, donde puedan ser alojados los documentos electrónicos (en algunos casos, LAOAP puede ofrecer servicios de alojamiento).
3. Capacidad para generar registros de metadatos conforme a las especificaciones de Dublin Core para cada uno de los documentos electrónicos de la colección (ver ejemplo en Anexo 2).

Para aquellas instituciones que ya cuentan con un acervo de documentos electrónicos, los pasos a seguir en el primer acápite son relativamente sencillos. Por ejemplo, si dispone de un conjunto de publicaciones en formato MS Word, lo mínimo que se tendría que hacer es convertirlos a PDF. En cambio, en el caso de las instituciones que cuentan únicamente con colecciones no digitales, por ejemplo fotografías o libros, el paso sería digitalizar o escanear, proceso que requiere de más tiempo y recursos. LAOAP puede facilitar asesoría y documentación sobre este proceso, pero no cuenta con fondos para la adquisición de equipos.

Para el segundo acápite, alojamiento en Web de los documentos electrónicos, es importante que los participantes en el LAOAP cuenten con un instrumento que les permita el alojamiento a largo plazo de sus documentos electrónicos en la Web. Para aquellos que ya manejan su propio servidor Web, únicamente se requiere espacio para el alojamiento de los archivos. Si cuentan además con un catálogo en línea, o algún sistema de manejo de contenidos en sus sitio, estos archivos pueden ser incorporados al esquema existente.

En el caso de aquellas instituciones que desean participar en el proyecto pero no tienen presencia propia en la Web, el LAOAP-LANIC puede proveer espacio en sus servidores durante un periodo interino para el alojamiento de los documentos y sus metadatos.

En tercer acápite, en cuanto a los metadatos existen varios escenarios posibles. Si ya cuentan con registros catalogados, por ejemplo en MicroIsis u otro paquete para bibliotecas, se espera que con la asistencia de personal del LAOAP, estos registros puedan ser “convertidos” o rehabilitados con las especificaciones mínimas de Dublin Core. Para los documentos no catalogados, el LAOAP proveerá una herramienta para el ingreso de datos en conformidad con las especificaciones requeridas.

Los metadatos de los recursos digitales de la colección tienen que ser también alojados en la Web de tal manera que puedan ser "cosechados" por el *Proveedor de Servicios*. Para esto, existen dos modalidades posibles. Como mínimo, se puede generar un archivo XML que reúna todos los metadatos de la colección, conforme a las especificaciones entregadas por el LAOAP. Para instituciones que cuentan con una plataforma tecnológica instalada de mayor

envergadura, se instalaría un Paquete de Proveedor de Datos en el servidor Web, que se encargaría de generar y disponer los metadatos bajo las especificaciones que se necesitan para ser "cosechados".

En síntesis, existen tres diferentes alternativas de posible ingreso al LAOAP como *Proveedor de Datos*.

1. Organizaciones que cuentan con documentos o publicaciones, pero sin presencia propia en la Web. En estos casos, LAOAP proveerá alojamiento temporal de los documentos y sus metadatos;
2. Grupos que tienen su propio servidor Web o acceso bajo algún esquema de alojamiento compartido, y cuentan con personal capacitado para el manejo de HTML y paginas Web;
3. Grupos que disponen de su propio servidor Web y de bases de datos u otras aplicaciones dinámicas en Web, y cuentan con personal especializado en programación y administración de sistemas.

En lo que se refiere a los metadatos, estas alternativas se perfilan de la siguiente manera:

Modelo	***Descripción***	***Perfil-Requisitos***
Básico	El *Proveedor de Datos* ingresa los metadatos por medio de un formulario de HTML facilitado por LAOAP. El formulario genera un archivo XML con los metadatos que luego pueden ser cosechados por medio de un SRG (Static Repository Gateway).	1. La colección no ha sido previamente catalogada. 2. *El Proveedor de Datos* no cuenta con personal capacitado en HTML.
Estático	El *Proveedor de Datos* genera un archivo XML que contiene los metadatos de todas las publicaciones en su colección. El *Proveedor de Servicios* instala un modulo SRG que puede cosechar los documentos XML colocados por los repositorios.	1. La colección consta de menos de 5 mil documentos. 2. Los registros de catalogación pueden ser exportados a, o incorporados en, un archivo XML. 3. El *Proveedor de Datos* tiene acceso a un servidor Web o a un servicio de alojamiento donde puede subir el archivo XML con los metadatos.
Interactivo	El *Proveedor de Datos* instala un "paquete puente" enlazado con su propio catálogo en línea. El	1. La colección ha sido previamente catalogada y el catálogo esta disponible en

	Proveedor de Servicios puede cosechar los metadatos directamente por medio del protocolo HTTP. No hay necesidad de modificar los registros del catálogo existente, ni de recatalogar los materiales.	línea. 2. El *Proveedor de Datos* cuenta con su propio servidor Web. 3. El *Proveedor de Datos* cuenta en su equipo de trabajo con un programador o un administrador de sistemas que puede confeccionar, configurar, y mantener el "paquete puente".

Para las alternativas *Estático* y *Básico*, el alojamiento de los documentos electrónicos y de los metadatos puede ser en servidor Web propio o en el servidor de LAOAP-LANIC.

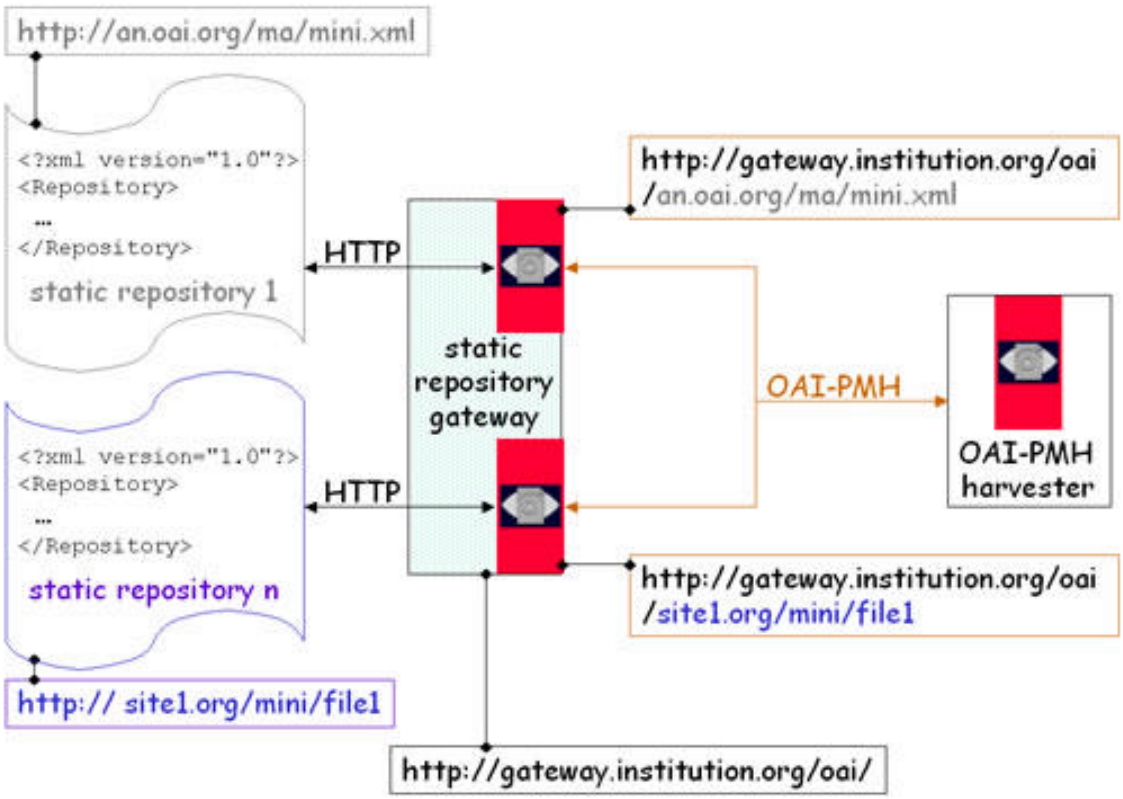
Si desea participar en el proyecto como *Proveedor de Datos*, comuníquese con nosotros con una carta que exprese su interés, y llenando el Formulario (ver Anexo 3) acerca del personal e infraestructura Web de su institución.

8. El *Static Repository Gateway*

Una de las novedades impulsadas en el marco de OAI por nuestro proyecto es el concepto de *Static Repository Gateway*, conocido por sus siglas en inglés como SRG. Esta alternativa, que ha sido reconocida formalmente como protocolo válido por parte de las autoridades que rigen la implementación de los sistemas OAI, surgió recientemente como una opción que hace más fácil la participación bajo el modelo OAI para potenciales proveedores de datos. Específicamente, en el modelo “tradicional” de OAI, para ser un proveedor de datos, se requiere instalar en el servidor remoto un software que responde interactivamente a pedidos hechos por la red por el proveedor de servicios. En la práctica, varios proveedores de servicios vieron que habían personas e instituciones interesados en participar en sus proyectos con sus publicaciones y metadatos, pero no tenían la capacidad para instalar, correr y mantener el software necesario.

El modelo SRG surge en ese contexto como una alternativa que permite al proveedor de datos exponer sus metadatos en forma de un archivo de XML “estático”, sin necesidad de instalar software adicional alguno. Esto abre la posibilidad de participar en este tipo de proyectos a un rango de instituciones mucho más amplio, al reducirse los requisitos mínimos para participar a simplemente tener archivos de texto completo, tener metadatos de dichos archivos, y de contar con espacio en Web para alojar los textos y el archivo XML con los metadatos (de hecho, la barrera es aún mas sencilla, pues para aquellos que no cuentan con presencia en Web, LANIC ofrece la posibilidad de alojar en sus servidores las publicaciones y los metadatos).

Los autores del SRG—Patrick Hochstenbach, Henry Jerez, y Herbert Van de Sompel—describen el modelo conceptualmente y visualmente de la siguiente manera:



Bajo esta modalidad, LANIC pretendería operar el SRG y el Cosechador (Harvester), y puede enlazarse con un número significativo de Proveedores de Datos montados cada uno como un archivo XML que contiene los metadatos. El SRG ha sido instalado y probado, y actualmente estamos trabajando en la generación de dos herramientas para facilitar la generación de metadatos y del archivo XML que sería un repositorio: una pequeña utilidad que tomaría un archivo con registros bibliográficos, por ejemplo exportado de MicroIsis, y lo convertiría a Dublin Core, y para los que no cuentan con los registros catalogados, un formulario en Web donde los metadatos pueden ser ingresados manualmente. En ambos casos, el resultado sería un archivo en XML con las etiquetas necesarias para que quede integrado como repositorio.

Referencias

Hochstenbach, Patrick, Henry Jerez, and Herbert Van de Sompel, The OAI-PMH Static Repository and Static Repository Gateway. Joint Conference on Digital Libraries, May 27-31, 2003, Houston.

Lagoze, Carl and Herbert Van de Sompel. The making of the Open Archives Initiative Protocol for Metadata Harvesting. 2003. Special issue of *Library Hi Tech* 2003, Volume 21, Number 2, Pages 118-128.

Open Archives Initiative, Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting - Specification for an OAI Static Repository and an OAI Static Repository Gateway, Protocol Version 2.0 of 2002-06-14, http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm

Open Archives Initiative, The Open Archives Initiative Protocol for Metadata Harvesting
Protocol Version 2.0 of 2002-06-14,
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Tennant, Roy, California Digital Library. Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers,
http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

Van de Sompel, Herbert, Jeffrey A. Young, and Thomas B. Hickey. Using the OAI-PMH ... differently. *D-Lib Magazine*, July/August 2003, v. 9, no. 7/8.

Anexo 1. La Biblioteca FLACSO-Chile

Junto con la creación de FLACSO en el año 1957 se establece, en Santiago de Chile, la Escuela Latinoamericana de Sociología - ELAS, gracias a un convenio entre la Universidad de Chile y UNESCO. Sus creadores tuvieron tres propósitos fundamentales al darle vida: formar profesores e investigadores a nivel de postgrado, realizar investigaciones sociológicas y asesorar técnicamente a aquellas instituciones que solicitaran su cooperación, viniendo a satisfacer una necesidad de los países latinoamericanos en su proceso de desarrollo. Con el propósito de apoyar las labores de docencia se crea la biblioteca de ELAS en el año 1958.

Ocho años más tarde, la FLACSO, aceptando la iniciativa y la ayuda financiera especial del Banco Interamericano de Desarrollo, crea la Escuela Latinoamericana de Ciencia Política y Administración Pública, ELACP. Desde su fundación, en el año 1966, se organiza una biblioteca especializada con el objetivo de permitir a los profesores, investigadores y estudiantes contar con material bibliográfico moderno y de alta calidad en materias de ciencia política, administración pública, teoría política, metodología y otras. Comienza sus labores, con la donación de 1.000 títulos pertenecientes a la biblioteca particular del primer Director de ELACP, señor Horacio Godoy.

A fines del año 1973 se inicia un proceso de cambio en la FLACSO, convirtiéndose Chile en Sede y estableciéndose la Secretaría General en Costa Rica. Como producto de estos cambios, las dos bibliotecas existentes se fusionan en un solo local, dando origen a la actual. Los duplicados, alrededor de 10.000 volúmenes, se envían a la reciente creada Sede de FLACSO en México, sirviendo de base a la futura Biblioteca Iberoamericana.

Evolución histórica de su fondo bibliográfico

1ª. Etapa 1957-1989: Catálogos manuales

2ª. Etapa 1990-1996: Inicio etapa de automatización (Proyecto UNESCO)

3ª. Etapa 1997-1999: Eliminación de catálogos manuales y puesta en marcha de un terminal de computador al público.

4ª. Etapa 1999-2002: Servicio electrónico y uso de bases de datos en CD para usuarios.

5ª. Etapa 2003- Catálogo Bibliográfico Público de Acceso en Línea

Recursos Bibliográficos

Nuestra Biblioteca es una de las más completas en el área de las ciencias sociales del país, contando en la actualidad, con una colección especializada que abarca sociología, ciencia política, relaciones internacionales, estudios de género, gestión pública, y otras.

Su fondo bibliográfico, incrementado en el transcurso de los años, constituye un depósito vivo y organizado no sólo de la historia social y política del país, sino también de todas aquellas obras clásicas tanto de sociología como de ciencia política, relaciones internacionales y afines.

Con el propósito de brindar servicio se cuenta con los siguientes recursos para los usuarios:

* Base de Datos Bibliográfica
* Catálogo en Línea
* Base de Datos de Prensa
* Base de Datos Fondo Documental Eugenio Ruiz-Tagle
* Base de Datos de Publicaciones Seriadas
* Base de Datos de Suscripciones
* Base de Datos de Canje
* Catálogo de Libros FLACSO-Chile 1980-1997
* Catálogo de Libros FLACSO-Chile 1998-2003
* Catálogo de Publicaciones Periódicas 2003
* Documentos de Trabajo 1975-1995: bibliografía analítica

Servicios

Para lograr sus objetivos, la biblioteca realiza diversas labores anexas a la de recopilar, analizar y sistematizar la información, las que están destinadas a dar respuesta a los requerimientos de información presentada por los usuarios.

Estas son:

* Referencia: en sala, telefónica, fax y correo electrónico, por medio de este servicio se localiza información concreta y se proporciona inmediata respuesta a la demanda de información presentada por el usuario.

* Servicios Bibliotecarios: préstamos en sala, a domicilio e interbibliotecarios con este servicio se ofrece en préstamo a los usuarios internos y externos, todo el material que conforma el fondo bibliográfico, a través de búsquedas automatizadas. Además se cuenta con los mecanismos para la obtención de material a nivel nacional o internacional, ya sea mediante préstamos interbibliotecarios y además canje, donación o compra en el caso de material extranjero.

* Perfiles Bibliográficos: listados o búsquedas de material en nuestra base de datos sobre temas específicos: este servicio ofrece información específica sobre un determinado tema, según el requerimiento del usuario. El servicio es proporcionado por medio de búsquedas automatizadas ya sea en nuestro catálogo en línea o en otras bases externas de información, cuando la ocasión lo requiera, para complementar y hacer más exhaustiva la búsqueda.

* Servicios de Alertas Bibliográficas: Con el fin de mantener al día a los investigadores, internos y externos y diseminar el material bibliográfico ingresado, biblioteca ofrece una serie de servicios:

    * Material Bibliográfico Ingresado a Base de Datos: es un boletín mensual que da cuenta de todo el material ingresado en el mes al Catálogo en Línea se compone de: libros, documentos, conferencias, analíticas de libros y revistas.

    * Catálogo de Publicaciones Periódicas: se confecciona con el fin de difundir y presentar a los centros de documentación la valiosa hemeroteca con que cuenta biblioteca. En este catálogo, ordenado alfabéticamente por título de publicación, se consignan los años de existencia, su ISSN, y su editor.

    * Catálogo de Publicaciones FLACSO-Chile: de periodicidad anual, tiene por objetivo informar al usuario sobre la producción bibliográfica intelectual de la Institución. Las referencias aparecen ordenadas por autor.

    * Catálogo de Documentos de Trabajo 1975-1995: tiene como objetivo dar a conocer las investigaciones y estudios realizados en FLACSO.

Estadísticas

Con el objetivo de mantener una relación de los trabajos que se realizan en biblioteca, se llevan a cabo diversas estadísticas, diarias y mensuales, las que anualmente sirven de base para el Informe Anual de Biblioteca que es presentado a la dirección de la Institución.

Anexo 2: Ficha de Metadatos del LAOAP

Latin American Open Archives Portal

Title:	La encuesta de opinión como vehículo de expresión política en contextos autoritarios: Un experimento metodológico
Creator:	Charlin, Marcelo
Subject:	Encuestas Encuestas de opinión pública Metodos de investigación social Técnicas de investigación social Análisis de datos Chile
Description:	Con el propósito de contribuir a la comprensión de los problemas relacionados con la interpretación de encuestas políticas, aplicadas en sociedades con alto nivel de polarización y bajo condiciones de represión política en contextos autoritarios, analizamos los datos provenientes de un experimento metodológico realizado para estudiar los sesgos que dicho tipo de encuestas puede presentar.
Publisher:	FLACSO-Programa Chile
Date:	1991-10
Type:	texto
Format:	text/pdf 79KB
Location:	http://lanic.utexas.edu/project/laoap/charlin.pdf
Language:	es
Relation:	EsParteDe Serie Estudios Políticos
Rights:	© 1991, FLACSO-Chile, Prohibida su reproducción.

Anexo 3: Formulario para Participantes

Formulario para Instituciones interesadas en integrarse al proyecto LAOAP

1. _ Tenemos presencia en Web en la dirección: http://_____
   _ No disponemos de presencia en Web, pero quisiéramos participar en el proyecto con alojamiento en LAOAP-LANIC.
2. _ Tenemos dominio propio en _____ y servidor propio con las siguientes características (facilitar una breve descripción del hardware y software del servidor).
   _ Tenemos dominio propio en _____ y nuestro sitio esta alojado con el proveedor de servicio _____.
3. Nuestro equipo de trabajo incluye:
   _ Webmaster con conocimiento de HTML
   _ Bibliotecario profesional
   _ Bibliotecario asistente
   _ Programador
   _ Administrador de bases de datos
   _ Administrador de sistemas

Si están interesados en participar en el proyecto LAOAP con publicaciones de texto, envíenos un resumen del contenido así como una estimación del total de publicaciones y el promedio de su extensión; qué porcentaje está en formato digital y qué porcentaje únicamente en papel; si cuentan o no con equipo y personal para escanear; y si las publicaciones han sido catalogadas y bajo que sistema.

Por favor, envíen una carta expresando su interés junto con este formulario a:

eloh@library.ucla.edu

Eudora I. Loh
Latin American and Iberian Bibliographer
Charles E. Young Research Library A1540Q
UCLA Box 951575
Los Angeles, CA 90095-1575 U.S.A.

Anexo 4: Glosario

---

* **AAU**: Association of American Universities

---

* **ARL**: Association of Research Libraries

---

* **CIRMA**: Centro de Investigaciones Regionales de Mesoamérica

---

* **Data provider**: (*Proveedor de Datos*)

  En el modelo OAI, el Proveedor de Datos se compromete a mantener en Web un conjunto de objetos digitales, respaldado por un conjunto de fichas de metadatos, ambos conforme a las especificaciones de OAI.

---

* **DC**: Dublin Core

  Formato de metadatos elaborado a partir de un amplio consenso internacional que cuenta con 15 campos descriptivos. Todos los campos son recomendados, ninguno es obligatorio. El formato DC es obligatorio para los proveedores de datos en el modelo OAI.

---

* **FLACSO**: Facultad Latinoamericana de Ciencias Sociales

---

* **Grey literature**: (*Literatura Gris*)

  Categoría de documentos que incluye documentos de trabajo, pre-impresiones, avances de investigación, y otros materiales difíciles de acceder y que generalmente son producidos fuera de los circuitos editoriales de carácter comercial.

---

* **GRN**: AAU/ARL Global Resources Network

---

* **Harvester**: (*Cosechador*)

  En el modelo OAI, el cosechador es una aplicación en el servidor que recupera los metadatos de los Proveedores de Datos por medio de pedidos interactivos formulados conforme las especificaciones OAI.

---

* **LANIC**: Latin American Network Information Center

---

* **LAOAP**: Latin American Open Archives Portal

---

* **LAPTOC**: Latin American Periodicals Tables of Contents

---

* **LARRP**: Latin Americanist Research Resources Project

---

* **OAI**: Open Archives Initiative (*Iniciativa de Archivos Abiertos*)

  La OAI es una iniciativa que busca desarrollar y promover estándares de interoperabilidad con el objetivo de facilitar la difusión de contenidos en la Web.

---

* **OCR**: Optical Character Recognition (*Reconocimiento Óptico de Caracteres*)

  El proceso de reconocer caracteres impresos o escritos por medio de un computador. Se requiere de software especializado para hacer OCR.

---

* **PDF**: Portable Document Format (*Formato de Documento Portátil*)

Un formato universal que preserva las fuentes, imágenes, gráficos, y diseño de cualquier documento original, independiente de la aplicación y plataforma utilizada para crearlo.

---

* **PMH**: Protocol for Metadata Harvesting (*Protocolo para Cosechar Metadatos*)

  El en modelo OAI, la PMH es un estándar sencillo para compartir metadatos entre diferentes sistemas.

---

* **Service Provider**: (*Proveedor de Servicios*)

  En el modelo OAI, el Proveedor de Servicios es el que opera el cosechador de metadatos para juntar los metadatos de los Proveedores de Datos, y luego ofrecer servicios de búsqueda avanzada para una comunidad especializada de usuarios.

---

* **SRG**: Static Repository Gateway

  En el modelo OAI, el SRG es una “aplicación puente” o intermediario instalado por un Proveedor de Servicios que permite que los Proveedores de Datos exponen sus metadatos en forma de un archivo plano en XML. En sistemas que cuentan con un SRG, los requisitos que necesita un potencial Proveedor de Datos son significativamente menos que en el modelo OAI “tradicional”.

---

* **TICFIA**: Technological Innovation and Cooperation for Foreign Information Access

---

* **TIFF**: Tagged Image File Format (*Formato de Archivo de Imagen con Etiquetas*)

  El formato de archivo estándar para gráficos de mapa de bits de alta resolución, utilizado generalmente por los programas de escaneo.