

Argentina

Los repositorios institucionales como vía de acceso abierto al conocimiento científico

.....
*Mariana Pichinini,
Biblioteca Prof. Guillermo Obiols
de la Facultad de Humanidades
y Ciencias de la Educación de la
Universidad Nacional de La Plata, Argentina*

Introducción

Para cumplir con las principales responsabilidades de universidades, la docencia e investigación, la divulgación y preservación de la producción científica constituyen pilares fundamentales del trabajo académico. Los procesos y métodos de difusión y preservación de dicho conocimiento son sujetos de cambio continuo: “En la actualidad, el avance de las tecnologías de la información y la comunicación ha ido variando el modo en que se divulga la información en las instituciones. La memoria académica de una universidad se define hoy en día como un repositorio que contiene los documentos generados por ella, de libre acceso y con permanencia en el tiempo. Las tendencias internacionales hacia el acceso abierto (*Open Access*) postulan la libre disponibilidad del texto completo de los documentos científicos a través de Internet para buscar, leer, descargar, distribuir, imprimir, indizar y cualquier otro propósito legítimo. Aparecen dos maneras de implementación: publicación del trabajo científico en una revista con acceso abierto y/o el autoarchivo por parte de los autores en repositorios institucionales [Berlin Declaration on Open Access 2003]” (Fushimi, Mallo, Pichinini 2005: p. 2).

El concepto de repositorio institucional está en parte ligado con la noción de biblioteca digital, esto es, recogiendo, almacenando, clasificando, catalogando, conservando, preservando y suministrando acceso a contenido digital, análogo con

las funciones convencionales de las bibliotecas respecto del material impreso.

La otra conexión de este concepto es con la noción de ‘interoperabilidad’ digital, relacionada con la *Open Archives Initiative* (OAI) y su *Protocol for Metadata Harvesting* (OAIPMH). Este protocolo permite la extracción normalizada de metadatos de los archivos y la recuperación de los mismos. Así, diferentes buscadores pueden ofrecer información de archivos separados y recuperarlos desde un sólo lugar.

Según un estudio realizado en 2008 por Carolina De Volder (De Volder 2008), en Argentina existen varias iniciativas de desarrollo de repositorios institucionales y/o bibliotecas digitales de acceso abierto, de los cuáles un 40% se trata de proyectos consolidados con inclusión de documentos en forma continua y un portal web estable, sobre un total de 15 proyectos reseñados.

La falta de políticas públicas y mandatos relacionados con el acceso abierto, sumado al poco apoyo económico y tecnológico, al desconocimiento y desconfianza de académicos e investigadores, a la falta de comunicación y cooperación entre los responsables de los proyectos y entre las autoridades de las instituciones, entre otros factores, son la causa del estado actual de avance en este tema.

Digitalización de la Memoria Académica de la Facultad de Humanidades y Ciencias de la Educación (FAHCE)

“La Biblioteca de la FAHCE de la Universidad Nacional de La Plata (BIBHUMA), Argentina, elaboró en 2005 un proyecto para la creación de la Memoria Académica de la Facultad, que propone la reunión y el registro sistematizado y permanente de su producción académica, plasmada en los documentos que se enumeran a continuación, en un único sitio web con acceso abierto y público:

- los trabajos (éditos o inéditos) de docentes e investigadores realizados en el ámbito de la FAHCE,
- las monografías aprobadas de alumnos de carreras de grado y posgrado que sean recomendadas para su pu-

blicación por los respectivos departamentos las tesis de doctorado y maestrías,

- las tesinas de licenciatura y trabajos de final de carrera,
- las publicaciones de todo tipo que edita la Facultad a través de su Comité Editorial, Departamentos Docentes, Centros de estudios e investigaciones y Cátedras, etc. los programas de materias, cursos y seminarios, así como los planes de estudio de
- las carreras de grado y posgrado que se cursan en su ámbito, la documentación de proyectos de investigación y extensión aprobados: proyectos, informes de resultados o avances, etc., proyectos e informes de becarios,
- las memorias de Departamentos Docentes, Centros de Estudios e Institutos de Investigación,
- toda otra documentación de carácter académico que se estime conveniente preservar y difundir” (Fushimi, Mallo, Pichinini 2005: p. 9).

Este proyecto se inició con una etapa prospectiva, con la inclusión de todos los documentos generados por la labor académica y científica de la institución a partir de 2006. Las etapas iniciales fueron:

- aprobación del proyecto y sus anexos normativos por el Honorable Consejo Académico de la Facultad
- adecuación de la normativa en reglamentos y procedimientos internos
- amplia difusión a toda la comunidad académica
- trabajo conjunto con las áreas involucradas para armonizar criterios y procedimientos
- designación de los recursos humanos solicitados y adquisición del equipamiento
- diseño y desarrollo de las colecciones digitales
- recopilación, obtención, procesamiento técnico y alta de los documentos

La etapa retrospectiva se fue encarando en función de requerimientos puntuales de preservación y acceso, como es

el caso de programas de estudio de asignaturas necesarios para estudios de historia de la educación en malas condiciones, publicaciones editadas por la Facultad de tirada agotada y mucho uso, documentos administrativos como convenios, resoluciones, normativa a las cuáles era importante acceder, etc.

Requerimientos al software

En el diseño y desarrollo de las colecciones digitales fue necesario establecer lineamientos y criterios a seguir para la manipulación y almacenamiento de los objetos digitales, por un lado, y la elección del software de gestión del proyecto y la biblioteca digital, por otro.

En cuanto a la preservación digital, es importante destacar que la mayor parte de los documentos pertenecientes a la producción actual de la Facultad son “nacidos digitales”, es decir, no requieren un proceso de digitalización previo para formar parte del repositorio.

En este sentido, se hizo hincapié en el formato de los archivos fuente, en la descripción de los mismos por medio de metadatos, y en la migración a un formato estable de almacenamiento a largo plazo.

En los casos en que se decidió digitalizar documentos impresos, se siguieron los estándares actuales recomendados en cuanto a formato de los archivos, estructura de almacenamiento, resolución espacial y profundidad de bits de las imágenes resultantes, proceso de OCR para obtención de texto, así como a los elementos descriptivos de preservación.

Teniendo en cuenta estos criterios, la elección del software debía cumplir con los siguientes requisitos:

1. Contar con algún mecanismo que permitiese almacenar y/o exportar los objetos digitales (metadatos más imagen/texto/audio/video) a un formato estable de archivo con fines de preservación utilizando lenguaje XML.
2. Incluir la descripción de los objetos digitales por medio de metadatos, la opción de definir esquemas pro-

pios así como también adoptar esquemas vigentes y realizar modificaciones sobre los mismos.

3. Permitir la indización y las búsquedas por el texto completo.
4. Manejar diferentes tipos de archivos como documento fuente (texto, imagen, video, audio).
5. Gestionar el depósito y administración de los objetos digitales.
6. Contar con soporte y documentación suficiente para capacitar a los recursos humanos existentes, con insuficiente formación informática de base.
7. Tratarse de un software libre, es decir, que pueda utilizarse para cualquier propósito, y modificarse para adaptarlo a necesidades puntuales.

La mayor parte del software libre que está presente en el mercado hoy día cumple con gran parte de ellos. El punto débil en casi todos es la gestión del repositorio o de la biblioteca digital, es decir, las herramientas y desarrollos disponibles para la manipulación de los documentos fuente, la interfaz de indización mediante metadatos y el sistema de uploading de recursos y preindización por parte del depositante.

Si bien sería ideal que un sistema brindase todas las características deseables, esta aspiración es reemplazada generalmente con la pretensión de que el sistema sea rápido y eficiente en gran parte de estos puntos, siempre y cuando permita la interoperabilidad con otros desarrollos que complementen las partes débiles.

En base a estas consideraciones, BIBHUMA decidió desarrollar el repositorio institucional utilizando dos programas:

1. Greenstone, un conjunto de programas de software diseñado para crear y distribuir colecciones digitales. En 2006 se comenzó a trabajar con la versión 2.70 para Windows como testeo y capacitación. Una vez completado el desarrollo de algunas colecciones y realizado el diseño del Portal web, se instaló la versión 2.74 para Linux en el servidor. En 2009 se migró a la versión 2.82 y se puso operativo el servidor OAI (Figura 1).

2. Plone, un Sistema de Gestión de Contenidos (CMS), utilizado para el Portal web de Memoria Académica. Se espera disponer de la interfaz de indización y upload de documentos por esta vía en breve (Figura 2).

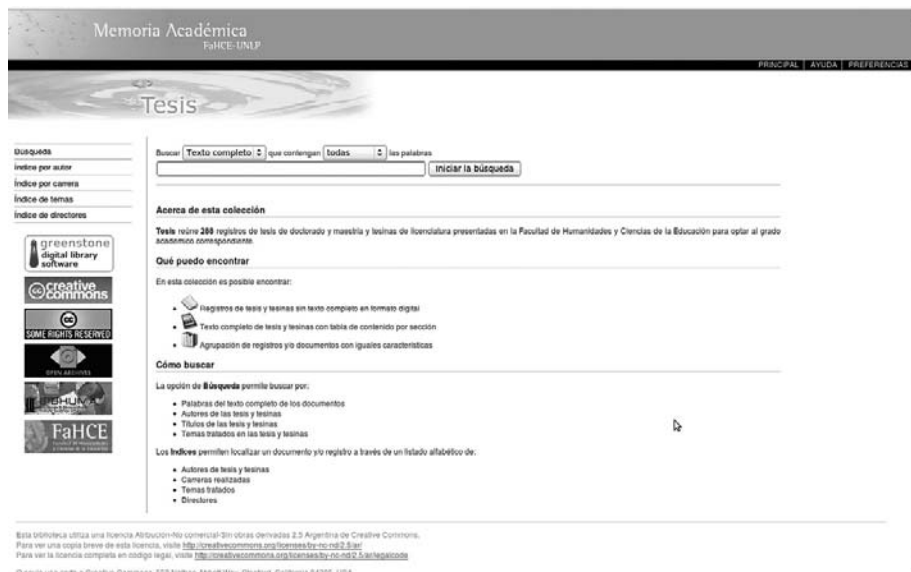


Figura 1. Colección de tesis desarrollada en Greenstone¹



Figura 2. Portal Memoria Académica desarrollado en Plone²

El flujo de trabajo para la indización por metadatos, procesamiento de los objetos digitales y preservación se realiza en las estaciones de trabajo. El depósito, la importación de los

documentos fuente y la generación de las colecciones se ejecutan directamente en el servidor.

El almacenamiento consiste en cinco puntos de acceso: Un acceso web para descarga de archivos .pdf (accesible mediante una referencia en el archivo de metadatos), un acceso web para la búsqueda y recuperación de documentos vía Greenstone, un acceso de trabajo para la importación de los objetos en Greenstone, un árbol de carpetas de trabajo por colección para el almacenamiento de los archivos resultantes de la manipulación y procesamiento de los objetos digitales así como un disco de resguardo externo con fines de *backup* y preservación.

Colección	Sólo Metadatos	Documentos a texto completo (incluye metadatos)
Planes de estudio		58
Programas de materias	5.410	1.003
Tesis de grado y posgrado	258	30
Artículos de revistas	2.688	1.083
Proyectos de investigación	482	
Convenios		50
Libros y capítulos de libros	1	2
Trabajos presentados a eventos	33	388
Normativa		11
TOTAL	8.872	2.625

Figura 3. Colecciones disponibles en Memoria Académica (datos noviembre 2009)

La cantidad de registros y textos completos visualizados de las colecciones existentes en Memoria Académica en 2009 fue 865.462, con un promedio de 96.000 descargas mensuales, según los registros de *log* del servidor web. La visibilidad alcanzada por el repositorio en el último año tuvo que ver básicamente con dos factores principales: El aumento en la canti-

dad de registros y documentos ofrecidos y la habilitación del servidor OAI a mediados de año (Pichinini 2009).

Funciones de Greenstone

El software Greenstone cuenta con un servidor OAI integrado, recién operativo en la versión 2.82, que permite a la biblioteca digital funcionar como un proveedor de datos (*data provider*), según lo especifica el protocolo OAIPMH 2.0. Esto significa que el repositorio/biblioteca digital expone mediante este protocolo sus metadatos para que las instituciones que proveen servicios de harvesting (*service provider*) cosechen, en forma automática y gradual, los registros de los objetos digitales disponibles en el repositorio.

El protocolo funciona de manera nativa con metadatos Dublin Core (DC), y el programa Greenstone provee un mecanismo de mapeo de cualquier esquema de metadatos utilizado en las colecciones a DC.

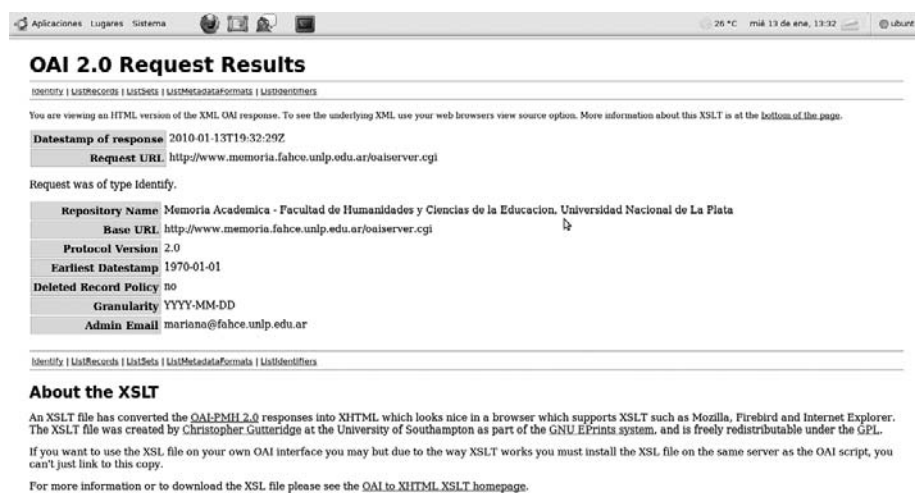


Figura 4. Resultado del verbo *Identify* sobre el servidor OAI de Greenstone³

Una vez puesto en funcionamiento este servidor, BIBHUMA registró a Memoria Académica como data provider en Open Archives Initiative⁴ y en otros registros de este tipo, como ROAR (*Registry of Open Access Repositories*), DOAR (*Directory of Open Access Repositories*) y University of Illinois OAIPMH Data Provider Registry.

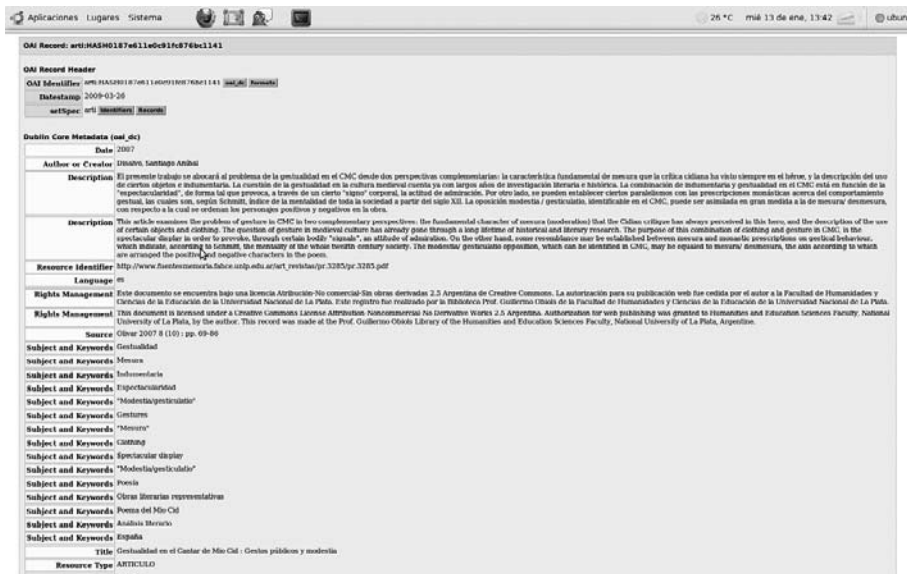


Figura 5. Resultado del verbo *GetRecord* sobre el servidor OAI de Greenstone

También se invitó a algunos de los service provider más conocidos a cosechar nuestros registros, así como OAISTER⁵ recientemente absorbido por WorldCat de OCLC, Biblioteca de Recursos de Aprendizaje UNIVERSIA⁶ y *Scientific Commons*⁷.

Ninguno de estos desarrollos habría sido posible sin la ayuda y el apoyo de la comunidad de usuarios de Greenstone en el mundo. Si bien este software creado por Ian Witten y su equipo de trabajo en la Universidad de Waikato en Nueva Zelanda brinda soporte permanente a las consultas que se realizan, hay un importante número de recursos disponibles para el neófito en la materia, así como para el desarrollador avanzado.

En el Portal de Greenstone se brinda acceso a toda la documentación existente sobre el software a través de un wiki que recibe aportes de muchos países. A su vez, existen dos listas de interés principales (de usuarios y de desarrolladores), en las que el 80% de las preguntas son respondidas por el equipo de Greenstone. También hay listas de interés en diferentes idiomas que dan soporte a diferentes comunidades de usuarios.

Formación del Centro Nacional de Greenstone

A través de la lista de usuarios en español⁸ se realizó una convocatoria para crear el Grupo de Promoción y Soporte de Greenstone para América Latina (GPSLA). La conformación de esta Red Regional de Greenstone para América Latina forma parte de las acciones que está llevando adelante la Universidad de Waikato para dar soporte y fortalecer la comunidad de usuarios y desarrolladores en países en vías de desarrollo.

Luego de un proceso de selección de organizaciones y mediante un convenio de cooperación, la Universidad de Waikato encargó a Prodigio Consultores⁹ la gestión de la creación de los centros nacionales que puedan encargarse de la promoción local del software. Con esta finalidad se convocó a la presentación de propuestas por cada país.

BIBHUMA, en forma conjunta con la Biblioteca Central Dr. Luis Federico Leloir de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, realizó una propuesta para postularse como Centro Nacional de Greenstone para Argentina, que fue aceptada.

La misma contempla cuatro aspectos específicos:

1. Relevamiento de la utilización del software en las universidades nacionales argentinas, que permitirá definir áreas de vacancia, realizar un informe del estado actual y crear un directorio de instituciones que utilizan Greenstone.
2. Actividades de capacitación abiertas a todo público (en función de los cupos disponibles, se dará prioridad a asistentes del ámbito académico), tomando como base el relevamiento realizado y enfocando la misma a los niveles introductorio, intermedio y avanzado.
3. Colaboración en el desarrollo de un sitio público en Greenstone¹⁰.
4. Traducción de documentación sobre Greenstone del idioma inglés al español.

El Centro se conformó efectivamente en septiembre de 2009 y realizó la primera reunión de usuarios de Greenstone en Rosario, Santa Fe, el 4 de noviembre de 2009, donde se pro-

bó un prototipo de encuesta que será lanzada oficialmente en febrero de 2010.

La conformación no solamente del repositorio sino también del sitio del Proyecto Memoria Académica de la Facultad de Humanidades y Ciencias de la Educación de la Universidad Nacional de La Plata conllevaron una serie de decisiones en cuanto a las políticas que esta Casa de Estudio debió seguir en su proyecto de conservar el patrimonio generado por su comunidad y ponerlo a disposición de aquellos que lo necesiten.

En tanto institución pública, la facultad entiende que tiene el derecho y el deber de dar a conocer de la manera más amplia posible los logros de sus integrantes. En este sentido, el registro en el sitio del proyecto de los resultados alcanzados por los mismos permite una mayor difusión de las actividades y publicaciones de la comunidad de la Facultad.

Las restricciones naturales de tiempo frente a un proyecto de tan gran envergadura han dado como resultado que las colecciones más provistas de documentos en formato digital sean aquellas cuyos documentos tienen índole más administrativa, y los artículos de revistas de las publicaciones editadas por la propia FaHCE. Si bien hasta el momento las tesis disponibles no son muchas, los doctorados de los últimos dos años dieron permisos con embargo, hasta tanto se concreten proyectos editoriales con empresas ajenas a la Facultad.

El impacto que el proyecto ha alcanzado en sólo dos años resultó completamente inesperado para los integrantes del plantel, Si bien la visibilidad del repositorio hacia afuera es alta, aún la comunidad académica de la Facultad no ha tomado como un objetivo propio la reunión y el registro de su producción académica por esta vía, y es el desafío que se plantea para los años próximos.

Referencias

Berlin Declaration on Open Access 2003:

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003. <http://oa.mpg.de/openaccessberlin/berlindeclaration.html> (verificado: 01/12/2009)

De Volder 2008:

De Volder, C.: Los repositorios de acceso abierto en Argentina. Situación actual. En: Información, Cultura y Sociedad, 2008, (19), p. 79-98. <http://www.scielo.org.ar/pdf/ics/n19/n19a05.pdf> (verificado: 25/02/2010)

Fushimi, Mallo, Pichinini 2005:

Fushimi, M.; Mallo, J.; Pichinini, M.: Memoria académica y científica: el rol de la biblioteca universitaria en la preservación y difusión del conocimiento generado en las universidades. En: Cuartas Jornadas de Sociología de la UNLP. La Plata 23-25 noviembre 2005. http://www.fuentesmemoria.fahce.unlp.edu.ar/trab_eventos/ev.518/ev.518.pdf (verificado: 25/02/2010)

Pichinini 2009:

Pichinini, M.: Implementación del protocolo para cosecha de metadatos OAIPMH 2.0. En: 7º Jornada de la biblioteca digital universitaria, Rosario 5-6 de noviembre 2009. <http://www.amicus.udesa.edu.ar/7bibliotecadigital/ponencias/PONENCIA%20PICHININI.pdf> (verificado: 25/02/2010)

Notas

¹ <http://www.memoria.fahce.unlp.edu.ar/librarysite=localhost&a=p&p=about&c=tesis&l=es&w=utf8>

² <http://www.memoria.fahce.unlp.edu.ar>

³ <http://www.memoria.fahce.unlp.edu.ar/oaiserver.cgi?verb=Identify>

⁴ <http://www.openarchives.org>

⁵ <http://www.oaister.org>

⁶ <http://biblioteca.universia.net/>

⁷ <http://www.scientificcommons.org/>

⁸ greenstone_es@freelists.org

⁹ <http://www.prodigioconsultores.com/>

¹⁰ <http://cng.fahce.unlp.edu.ar/>