

Satellite Meeting: *Up Up and Away: High density storage for library materials*
10 – 11 August 2016
Library of Congress, Washington DC, USA

Repositorio digital para la preservación de la investigación científica. Caso de estudio: Facultad de Comunicación de la Universidad de la Habana

Alejandro Medina González

Departamento de Ciencias de la Información de la Facultad de Comunicación de la
Universidad de la Habana

Havana, Cuba

E-mail address: ale.medina.glez86@gmail.com



Copyright © 2016 by Alejandro Medina González. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Resumen:

A partir del año 2014 la biblioteca de la Facultad de Comunicación de la Universidad de la Habana detectó una crisis de conservación, preservación, almacenamiento, búsqueda y recuperación. El problema se debe a la gran cantidad de tesis impresas y en soportes digitales que se deterioraban, además del aumento continuo del ingreso de nuevas tesis y de producción de materiales audiovisuales. Todo ello se considera un problema para la salvaguarda del patrimonio documental científico investigativo de la institución. Por esta razón se propuso la implementación de un proyecto de Repositorio Digital. En tal sentido este trabajo describe las etapas del diseño, implementación y evaluación del Repositorio Digital. Además expone como buenas prácticas el empleo de estudiantes de la carrera de Ciencias de la Información asociados a todas las etapas del proyecto y el Kow How desarrollado en un entorno con pocos recursos financieros. Finalmente, se exponen brevemente algunos resultados alcanzados en la implementación del repositorio, su proyección futura, su funcionamiento en acceso abierto para los próximos dos años y su contribución a la preservación del patrimonio documental científico de la Facultad de Comunicación.

Palabras clave: Repositorio Digital, Almacenamiento Masivo, Preservación Digital

Introducción

La Facultad de Comunicación de la Universidad de la Habana, cuenta con una pequeña biblioteca especializada con un relevante patrimonio investigativo documental en Ciencias de la Información y de la Comunicación, que incluye resultados investigativos en tópicos como: publicidad, marketing, gestión organizacional, comunicación institucional, periodismo hipermedia, periodismo impreso, documentales, bibliotecología, estudios métricos, archivística, gestión de la información, documental y del conocimiento, entre otros. Estas

investigaciones representan la evolución y desarrollo de una parte del sector informacional en Cuba, y en alguna medida, sirven de apoyo a la memoria institucional.

Principalmente, la biblioteca almacena y pone a disposición de sus usuarios tesis de grado, de maestría, y de doctorado; así como materiales audiovisuales, conferencias, libros, revistas y artículos de revistas, entre otras tipologías documentales que sustentan el fondo bibliográfico especializado. En esencia, la biblioteca representa no sólo la información contenida en sus documentos, sino un valor superior que tiene como propósito salvaguardar y difundir la memoria investigativa de un capital intelectual que narra y describe prácticas de una de las instituciones académicas más prestigiosas de Cuba.

Al igual que muchas instituciones afines, los retos a los que se enfrenta la biblioteca de la Facultad de Comunicación son el tiempo, el espacio y la escasez de recursos materiales para la conservación y la difusión. Los principales documentos afectados, debido a los factores antes mencionados, son las tesis y los materiales audiovisuales. En tal sentido, existe un deterioro de las tesis en formato impreso, contaminadas por las condiciones ambientales. Se almacenan además en el fondo de la biblioteca, cajas que contienen materiales de los cuales se desconoce el contenido y/o forma de los documentos. En esta misma situación, se encuentra información almacenada en discos de 3¼, pendiente por recuperar, pero sólo diagnosticable a través de una tecnología obsoleta.

Un elemento importante a considerar es la alta productividad de investigaciones que tienen como resultado materiales audiovisuales, que se entregan a la biblioteca en CD o DVD, que no se procesan de forma adecuada, ni se realizan acciones para preservar su contenido.

Atendiendo a los problemas identificados, se propuso como solución inicial, una estrategia orientada a la preservación de la información y del conocimiento, contenidos en los fondos de la biblioteca, y al acceso a estos recursos de forma eficiente; que toma como base la implementación de un Repositorio Digital, como herramienta tecnológica de soporte técnico. En tal sentido, el propósito de este estudio es presentar los avances realizados por la biblioteca de la Facultad de Comunicación, desde el diseño hasta una breve evaluación de su Repositorio Digital después de dos años.

Repositorio digital para la preservación y difusión de la producción científica de la Facultad de Comunicación de la Universidad de la Habana.

Los repositorios institucionales pueden ser definidos desde distintas perspectivas manteniendo su esencia. Las definiciones encontradas en la bibliografía consultada giran alrededor de criterios como: el contexto en el que se desarrolla, las características que definen estos contextos, y las formas genéricas que pueden tener un carácter tecnológico o pueden estar determinadas por los contenidos que almacena.

De manera general Crow, SPARC y la *Association of Research Libraries* (2002), definen un repositorio digital, como una colección digital soportada por un sistema electrónico que captura, preserva, y provee acceso a los productos de trabajo digitales de una comunidad, que puede pertenecer a una institución académica como una universidad, y representa la salida intelectual de una o múltiples instituciones. Además se caracteriza por ser:

- Definido institucionalmente: debido a la captura de la investigación original y otras propiedades intelectuales generadas por la institución. De esta forma, un repositorio institucional representa la personificación histórica y tangible de la vida intelectual, y las

salidas de la institución. El contenido puede ser depositado por el creador, propietario o un tercero en su nombre

- Académico: dependiendo de las metas establecidas por cada institución, un repositorio institucional puede contar con cualquier producto generado por los estudiantes de la institución, facultativos, investigadores no facultativos, u otro personal.
- Acumulativo y perpetuo: la aceptación de productos es constante y procura establecer una normativa para el mantenimiento de estos en el tiempo.
- Abierto e interoperable: provee a la institución de pocas o ningunas barreras a los productos intelectuales generados por la institución incrementando la concientización hacia la contribución de los investigadores y el acceso de las comunidades. Los sistemas de repositorios institucionales deben ser interoperables para proveer por múltiples vías a los motores de búsqueda y otras herramientas de descubrimiento.

Metodología del proyecto

Atendiendo al contexto expuesto y a las concepciones existentes sobre repositorios digitales, se analizaron dos metodologías: una para el diseño del proyecto de repositorio, tomando la propuesta de Barton y Water (2004); y la otra, de Esteban Navarro (2006), para el diseño del repositorio como un servicio de información digital de la biblioteca. Las etapas se definieron de forma tal que pudieran alcanzarse escalablemente en el tiempo, admitiendo la incorporación de recursos con los que inicialmente no se contaba. Las fases establecidas para el proyecto fueron: Planificación, Diseño, Implementación y Mantenimiento, con un componente de evaluación recursivo durante cada una de las etapas, que incluyó la documentación y supervisión del cumplimiento de lo planificado. De esta forma, quedaron las siguientes, con sus respectivas salidas parciales:

- Planificación:
 - Responsable del proyecto
 - Cronograma de trabajo
 - Identificación de requisitos
 - Disponibilidad de los recursos
 - Análisis de la plataforma de software a emplear
 - Equipo de trabajo (Roles y funciones)
- Diseño:
 - Justificación del Repositorio
 - Definición del Repositorio: misión, objetivos, metas y funciones
 - Alcance
 - Tipología Documental
 - Proveedores de información
 - Flujos de Trabajo
 - Servicios
 - Estructura organizativa de las colecciones
 - Metadatos
 - Arquitectura de Información
 - Políticas o pautas del repositorio
 - Difusión y promoción del repositorio

- Implementación:
 - Instalación del Software
 - Personalización
 - Alfabetización informacional

- Mantenimiento:
 - Migración del software
 - Revisión de las políticas
 - Auditoria de contenidos

Limitaciones

A priori, el proyecto resultó muy difícil de ejecutar, debido a la ausencia de financiamiento, escasez de recursos materiales, dificultades tecnológicas, problemas de procesamiento de la información y de la documentación de los fondos (malas prácticas o desestructuración de los documentos), y ausencia de personal en todos los roles necesarios (dígase bibliotecarios y programadores). Todo esto, fundamentalmente, influyó de forma negativa en los tiempos planificados y en las capacidades de procesamiento y rendimiento del Repositorio.

Pautas para el uso y mantenimiento del repositorio

Una parte importante para la disponibilidad, funcionamiento y utilidad del Repositorio Digital, parte de la regulación y estandarización de procedimientos. El propósito fue buscar alguna normativa inmediata, que abarcara el accionar acerca de aspectos como el empleo de esquemas de metadatos, formatos de depósito de archivos digitales, medidas de preservación digital, y acceso a la información digital.

La creación de estas pautas está basada en un proceso de adecuación de un conjunto de políticas de repositorios digitales de acceso abierto en línea. Los resultados se presumieron que tendrían un carácter temporal, hasta que quedase constituida la política general universitaria para repositorios digitales de la red de bibliotecas de la Universidad de la Habana.

Las pautas tuvieron un basamento experiencial, al realizarse una búsqueda en la web para resumir las políticas de repositorios homólogos con carácter institucional y/o universitario académico. Se pudieron consultar los sitios web asociados a los repositorios de las siguientes instituciones:

- Bibliotecas de la Universidad de Texas en Austin (Estados Unidos)
- Consejo Superior de Investigaciones Científicas “Digital CSIC” (España)
- Universidad Andrés Bello (Chile)
- Universidad de Arizona (Estados Unidos)
- Universidad de Cardiff (Reino Unido)
- Universidad de Exeter (Reino Unido)
- Universidad de Granada “DIGIBUG” (España)
- Universidad de Pittsburg “D-scholarship” (Estados Unidos)
- Universidad de Salamanca “Gredos” (España)
- Universidad de Ulster (Reino Unido)

- Universidad Edge Hill (Reino Unido)
- Universidad Nacional de La Plata “SEDICI” (Argentina)
- Universidad Tecnológica de Delf (Holanda)

En la mayoría de los casos, se encontraron aspectos en sus políticas relacionados con el contenido, los metadatos, el acceso a los documentos, el depósito, la preservación, los derechos de autor sobre la propiedad intelectual, la difusión de los servicios del repositorio y las estadísticas que se generan. Sobre esta base, se establecieron pautas sobre:

- 1) Los contenidos
- 2) Los metadatos
- 3) El acceso
- 4) El depósito (incluye materiales especiales como: imágenes, videos y audio)
- 5) El autoarchivo
- 6) La migración
- 7) El expurgo
- 8) El formato de los documentos
- 9) Las condiciones de respeto al derecho de autor y la propiedad intelectual (incluye un documento normativo, con las licencias *Creative Commons* que se otorgan)
- 10) La difusión de los servicios

De esta forma, las pautas diseñadas para el uso y mantenimiento del Repositorio Digital, constituyen un instrumento para normalizar las actividades, con el propósito de suplir inicialmente la carencia de una política que regule los elementos mencionados.

Desarrollo del Repositorio Digital

El punto de partida para comenzar el desarrollo del repositorio, fue establecer como objetivo general: *Crear un servicio digital para almacenar, preservar y difundir la información y el conocimiento de las prácticas investigativas de la Facultad de Comunicación de la Universidad de la Habana.*

En tal sentido, el repositorio se diseñó con dos comunidades: una de Tesis y otra de Producción científica, de profesores y estudiantes:

- La comunidad de Tesis, contiene las colecciones: Ciencias de la Información, Comunicación Social y Periodismo, en correspondencia con las especialidades que se imparten en la Facultad. Dentro de cada una de ellas, se establecieron como sub-colecciones: Pregrado, Maestría y Doctorado, en correspondencia con los niveles de docencia que se imparten, y con la investigación asociada que se realiza.
- La comunidad Producción científica, contiene artículos científicos, participaciones en eventos, en congresos, paneles y otros en los que participen los profesores y estudiantes.

Los servicios que ofrece el repositorio digital, se basan en el almacenamiento, preservación y difusión de información y de conocimiento, resultado de la producción científica de la facultad. Otros servicios específicos incluyen reportes de estadísticas, acceso a los metadatos y descarga del documento a texto completo (de momento) desde la red interna de la institución.

Desde el punto de vista organizativo, el equipo de trabajo para el procesamiento se conformó con estudiantes, profesores y bibliotecarios vinculados a asignaturas docentes donde se estudia análisis documental (catalogación, indización y clasificación), con el propósito de normalizar la documentación. De esta forma, la docencia se beneficia del repositorio, y viceversa, al usarlo como laboratorio para los estudiantes y profesores durante el ejercicio de sus clases y en el periodo de prácticas laborales, donde los estudiantes que estudian Ciencias de la Información, se insertan en la biblioteca de la Facultad y cumplen con las tareas asignadas.

El plan de acciones para la ejecución e implementación del repositorio, priorizó la carga masiva de las tesis, pues el ritmo de producción de éstas es cada vez más grande debido al aumento de las capacidades de matrícula. Al respecto, fue preciso comenzar con un proceso de normalización documental de forma. Un grupo de estudiantes asociado al proyecto de desarrollo de la biblioteca, se encargó de estructurar los pedazos de archivo en los que se encontraban divididos los documentos digitales, e integrarlos en un único archivo. Esta situación se producía anteriormente, debido a la carencia de normas para estandarizar el formato de entrega de los documentos al ser depositados. Por ejemplo: una tesis se entregaba digitalmente no como un solo archivo, sino en un folder con secciones en diferentes archivos como la portada, tabla de contenidos, capítulos, anexos, referencias, entre otros; cada uno de ellos era un archivo sin vínculos entre sí, representando cada uno de ellos documentos separados, lo cual afectaba la integridad de la tesis y su autoría.

Ante los problemas de normalización documental encontrados, se establecieron las pautas antes mencionadas y, en particular, se diseñó una plantilla para uniformar la composición estructural documental y de los metadatos de las tesis de todas las categorías científicas. Bajo la actual estricta política de la Facultad y de la biblioteca de cumplir con el formato establecido, en cada entrega de tesis se comprueba dicha plantilla.

Por esta razón, el orden de procesamiento para crear las colecciones e ingresar los documentos al repositorio dio prioridad a las tesis de doctorado en primer lugar, después a las de maestría, y por último a las de pregrado. Para un segundo momento, se incluyó la recopilación de los artículos de revista, participaciones en eventos y congresos, y otros materiales que pueden contemplarse.

Se estableció además que, al menos en los comienzos del repositorio, no se emplearía el autoarchivo, sino que se haría a través de terceros. En este caso, la biblioteca tiene el rol de supervisora del proceso y los estudiantes que trabajan con la biblioteca, el de procesadores de los recursos. Esta asignación de roles y funciones, fue necesaria para definir el flujo de trabajo de publicación de metadatos y documentos en el repositorio.

Finalmente, otro grupo de estudiantes trabaja en la búsqueda y recuperación de los artículos de revista, participaciones en eventos, congresos y otros materiales, de los cuales la facultad tiene referencia en los últimos 5 años, pero no se conserva en la biblioteca el documento a texto completo. En muchos casos, la solución ha sido solicitarles los documentos a los autores. En el caso de los profesores e investigadores que tienen un vínculo directo con la Facultad resulta más fácil su recuperación, no siendo así para aquellos que ya no lo hacen. De cualquier forma, el trabajo de recopilación se concentra en recuperar los recursos de información producidos en los últimos 15 años, empleando mayoritariamente la web para su localización.

Implantación del software

Mediante un análisis previo de la plataforma tecnológica, se decidió emplear *DSpace*, por ser un software de libre y de código abierto, sustentado en una comunidad desarrolladora, lo que permite incorporar avances sin muchas condiciones informáticas y de programación. Asimismo, *DSpace* cumple con estándares para la interoperabilidad con otros sistemas.

Para comenzar, se instaló la versión 1.8.3 de *DSpace*, aunque recientemente con la adquisición de nuevos servidores que aumentan las capacidades de almacenamiento (anteriormente de 100 GB ahora de 1TB con posibilidad de 2TB) se ha migrado a la versión 4.x.

Una vez instalado, el *DSpace* se procedió a la personalización de los metadatos atendiendo a las características de las colecciones y a la política de procesamiento de la biblioteca (empleando como herramienta para la descripción catalográfica, las Reglas de Catalogación Angloamericanas), empleando el esquema de metadatos *Dublin core* con pequeñas adecuaciones, donde los campos indispensables para las tesis son:

- Autor (es)
- Título
- Resumen
- Palabras Clave
- Año (año de defensa de la tesis)
- Tutor(es)
- Tipo de tesis

En tanto para los otros recursos se propusieron como indispensables:

- Autor
- Título
- Resumen
- Palabras Clave
- Año
- Fuente
- Descripción

Actualidad del Repositorio Digital de la Facultad de Comunicación

Existe un proyecto que llevan a cabo el Departamento de Ciencias de la Información y el Vicedecanato de Información y Desarrollo Tecnológico de la Facultad de Comunicación, para el desarrollo de la biblioteca, en el cual se incluye el Repositorio Digital como elemento principal. Entre sus primeros resultados se aprecia la propuesta de directrices para alcanzar y promover el acceso abierto, pautas para el depósito, preservación, el procesamiento y acceso, y la adopción y validación de licencias *Creative Commons*. Se pretende que, al término de este año, el repositorio pueda suscribirse en el Registro de Repositorios de Acceso Abierto (ROAR por sus siglas en inglés).

Otras actividades que se realizan son:

- 1) Auditorias al repositorio digital, priorizando al procesamiento, el control de los flujos de trabajos y el análisis de estadísticas.

- 2) Acciones de preservación, que se basan fundamentalmente en la creación de *backups* o salvas de la bases de datos del repositorio. Copias mensuales de los servidores y al término y mediados de semestre, por ser los dos momentos del año donde más documentos ingresan a la biblioteca.
- 3) Un plan de Alfabetización Informacional, que ha comenzado este curso, con la impartición de conferencias talleres a los estudiantes de primer año acerca de las funciones y usos que pueden hacer del repositorio para mejorar el aprovechamiento del mismo. Por otro lado, se preparan acciones para la promoción orientadas a los profesores, donde se expresa la necesidad de contribuir con el uso y difusión del repositorio. También se contempla la necesidad de realizar talleres sobre acceso abierto y colaboración científica, para fomentar una cultura organizacional sustentable.
- 4) Preparación de un plan para la digitalización de las tesis que se encuentran en formato impreso, comenzando por las más deterioradas (en este punto aún no se dispone de la tecnología pertinente).
- 5) Creación de un listado para el control de autoridades y su automatización, así como la actualización y digitalización de un listado de materias para la descripción de los recursos de información.

Entre los resultados visibles, que aparecen al término de dos cursos de estar funcionando el repositorio digital, se encuentra el acceso incrementado a las tesis defendidas entre los años 2010 y 2016, cuya cantidad asciende a un total de 1568, sumando todas las modalidades y categorías científicas (pregrado y posgrado). En caso de las de maestría y doctorado, se encuentran todas las tesis defendidas desde el año 2000.

Además, se ha subido al repositorio, un promedio de 120 tesis por año en estos dos últimos cursos, y ha disminuido el tiempo en el que se procesan las mismas. Se continúa el ingreso de las tesis anteriores al año 2004; y se ha logrado reunir un total de 226 publicaciones de profesores de los últimos 15 años, que podrían formar parte del repositorio al término de julio del presente año. Parte de las acciones para recuperar los artículos y otros resultados de investigación de los profesores y estudiantes, es localizarlos en la web y solicitárselos a los propios autores.

Conclusiones

Los Repositorios Digitales constituyen una valiosa herramienta para las bibliotecas universitarias. Un paso importante en la funcionalidad y organicidad de un Repositorio Digital es la normalización, en especial, de los metadatos y estructuras formales de los contenidos de los documentos que suelen ser de distintas tipologías. En tal sentido, los bibliotecarios tienen un rol importante como gestores de contenidos, de colecciones y promotores de los servicios del repositorio.

El Repositorio Digital de la Facultad de Comunicación de la Universidad de la Habana se crea con la misión de aumentar el acceso a la información y el conocimiento utilizando las condiciones tecnológicas de infraestructura existente. El repositorio además posibilita almacenar, compartir, difundir y preservar los recursos de información para la investigación y

la docencia. Para su desarrollo y mantenimiento futuro se requiere de la participación de profesores, informáticos, bibliotecarios y estudiantes a fin de beneficiar a la comunidad.

En su conjunto, las pautas para el uso y mantenimiento del repositorio han sido esenciales para que, tanto usuarios como administradores del repositorio, tengan conocimiento de lo que se puede o no hacer. Alguno de los aspectos que se definen, se centran en las tipologías documentales que se aceptan en el repositorio, el esquema de metadatos que se emplea, así como los metadatos de descripción mínimos que deben tener los recursos a depositar. Por otro lado, están los niveles de acceso que se asignan a las colecciones, quiénes depositan y cómo pueden hacerlo. Se definen también algunas medidas para la preservación del repositorio y cómo respetar la propiedad intelectual a través del reconocimiento de los autores sobre su obra. Las pautas de disseminación de los servicios, propone la utilización de todos los espacios con que cuenta la Facultad de Comunicación para promover el uso del repositorio.

Reconocimientos

Este artículo ha sido posible por el esfuerzo colaborativo de muchos estudiantes de Ciencias de la Información de la Facultad de Comunicación de la Universidad de la Habana. Primero, me gustaría agradecer al Consejo de Dirección de la institución, en especial a la Dra.C Déborah Torres Ponjuán Vicedecana de Información y Desarrollo Tecnológico, por el apoyo ofrecido en cuanto a la promoción de un cambio cultural respecto a la aceptación del repositorio y el cumplimiento de las normas establecidas para su utilización. También, me gustaría agradecer a los profesores y estudiantes de Ciencias de la Información asociados al proyecto de desarrollo de la biblioteca, y a las bibliotecarias, por su aporte durante el proceso de normalización. Además, es preciso reconocer al grupo de Tecnologías de Información y Comunicación de la Facultad de Comunicación por el apoyo técnico en la instalación, migración y otros aspectos de mantenimiento, así como al señor Marc Goovaert, director de la biblioteca de la Universidad de Hasselt (Bélgica), quien ha colaborado desinteresadamente y puesto a disposición del autor de este trabajo su conocimiento, experiencia y ayuda técnica para la migración y personalización de *DSpace*.

Referencias

- Crow, R., SPARC (Organization) y Association of Research Libraries. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, D.C: SPARC.
- Esteban Navarro, M. Á. (2006). Planificación, diseño y desarrollo de servicios de información digital. En J. Tramullas y P. Garrido (Coords.), *Software libre paraservicios de información digital*. (pp. 23-43). Madrid: Pearson Prentice Hall.
- Barton, M. R. y Waters, M. M. (2004). *Cómo crear un repositorio institucional. Manual LEADIRS II* (pp. 169). The Cambridge-MIT Institute (CMI).
- Kopper, D. N., Vargas, M. J., & American Library Association. (1983). *Reglas de catalogaciónangloamericanas*. Washington, D.C: Organización de los Estados Americanos.