

Repositorio Institucional de la PUCP

Lecciones Aprendidas

Mg. Jorge Alberto Solís Tovar
Pontificia Universidad Católica del Perú

Resumen:

El hombre en su afán de ampliar sus conocimientos siempre ha tenido la necesidad de comunicarse y preservar el conocimiento adquirido. Para ello ha recurrido a diversos medios que la tecnología de cada época le proporcionaba: el papiro, los manuscritos, los impresos, y últimamente, la tecnología digital y multimedia.

Los repositorios son las formas de almacenar los conocimientos en esta nueva tecnología digital y su auge se ha disparado con el incremento exponencial en la capacidad de almacenamiento y velocidad de recuperación de los contenidos que ofrecen los dispositivos digitales y la posibilidad de acceder a la información, desde cualquier punto del planeta, gracias a las redes de comunicaciones.

La PUCP al igual que muchas otras instituciones, ha encontrado en los Repositorios Digitales un medio eficaz y eficiente para divulgar el conocimiento que genera y la cultura e historia que alberga, por lo que desde el 2010 ha desarrollado un plan gradual de implementación de un Repositorio Institucional.

Desarrollo:

1. La Generación del Conocimiento.

Información es resolver una incertidumbre. Conocimiento es el resultado de la información y la experiencia. Es algo que se acumula en el tiempo, partiendo de los conocimientos anteriores e incluso modificándolos cuando la experiencia demuestra que estaban errados. Para ello tanto la información como el conocimiento deben poder ser transmitidos y almacenados.

En la era prehistórica la transmisión era oral, a distancias muy cortas y sin posibilidad de almacenamiento, salvo en la mente de los receptores de la información.

Por ello un primer avance en este sentido fue la aparición de la escritura y el papel (antes los papiros), que permitieron almacenar dicho conocimiento en impresos, lo que da origen a la aparición de las Bibliotecas para almacenarlos.

Un segundo gran impulso fue la invención de la imprenta, por Johannes Gutenberg, a mediados del siglo XV, pues con la imprenta se podían obtener muchas copias iguales de cada impreso, que se almacenaban en bibliotecas ubicadas en muchas partes del planeta (incluyendo América) acelerando así la difusión del conocimiento y su preservación.

2. La Era Digital.

Un tercer hito importante se produce a mediados del siglo pasado, con la aparición de las computadoras, máquinas de procesar información en base a la tecnología digital, es decir aquella que permite representar cualquier información en base a los dígitos 1 (uno) y 0 (cero).

En sus inicios la capacidad de almacenamiento y velocidad de proceso de los computadores eran muy limitados, por lo que estos solo se usaban para procesar datos alfanuméricos.

Sin embargo, en los algo más de 70 años de existencia del computador, estas capacidades se han incrementado a un ritmo exponencial, lo que ahora les permite almacenar no solo información alfanumérica, sino también otras expresiones humanas capaces de ser digitalizadas (convertidas a un conjunto de ceros y unos) como documentos impresos, fotografías, audios, videos y películas.

3. Los Repositorios Digitales.

La información alfanumérica se almacenaba en archivos planos, en los que cada registro era identificado por una o más claves que permitían su recuperación. Los archivos se usaban en procesos transaccionales, principalmente de índole administrativo (contabilidad, facturación, logística, planillas, etc.) aunque también en otros campos, como el catálogo de las bibliotecas, donde reemplazaban con ventaja a las antiguas fichas mecanografiadas que se usaban para la búsqueda de los libros existentes en la biblioteca.

Posteriormente, al aumentar las capacidades del computador, los diferentes archivos planos relacionados con un mismo tema se consolidaban y agrupaban en Bases de Datos, en los cuales los campos de información seguían siendo alfanuméricos y la forma de acceder a ellos era mediante múltiples claves de acceso.

El crecimiento vertiginoso de las capacidades de almacenamiento, y de la velocidad de proceso de los computadores, permitió, en años relativamente recientes, incluir en los registros de las Bases de Datos, campos de gran amplitud, que podían contener no solo información alfanumérica, sino también documentos impresos, fotografías, audios, películas y videos, todos ellos debidamente digitalizados, los cuales convivían con los campos de información alfanumérica de los archivos tradicionales.

En otras palabras, estos registros multimedia ya no albergaban únicamente información transaccional, sino que albergaban **conocimientos**, y si bien tenían la estructura de las Bases de Datos tradicionales, requerían otras formas de acceder a ellos. Por este motivo se reemplazaron claves alfanuméricas de acceso, por los denominados **metadatos**, que son identificadores que permiten acceder al conocimiento contenida los repositorios.

4. La importancia de las comunicaciones.

Estos avances en el almacenamiento y preservación del conocimiento no hubieran tenido el mismo impacto, si, paralelamente, no se hubieran desarrollado las comunicaciones, gracias a los satélites artificiales, que permiten que las ondas emitidas en un punto del planeta puedan ser visualizadas o recibidas en cualquier otro punto del planeta.

De esta manera se reemplazaron las líneas físicas de comunicación por sistemas inalámbricos, de gran capacidad de transporte de información, lo que permite que se pueda acceder a un repositorio desde cualquier punto del planeta en el que se tenga acceso a esta red de comunicaciones (internet).

5. Acceso Abierto.

Tradicionalmente cada persona tenía que comprar los libros que quería leer o asistir a una biblioteca pública que lo tuviera para leerlo en la sala de lectura. En algunos casos (bibliotecas

institucionales) las bibliotecas prestaban a sus usuarios registrados, bajo ciertas condiciones, los libros que requerían.

Esta realidad está cambiando aceleradamente con la producción de libros digitales y la digitalización progresiva de los libros impresos existentes en las bibliotecas. Las casas editoriales o que se dedican a este rubro, han armado grandes repositorios de libros digitalizados, a los cuales se puede acceder, afiliándose, previo pago, como usuario del respectivo repositorio.

Normalmente son las bibliotecas las que se afilian a estos repositorios de libros, y mediante esta afiliación le dan el servicio a toda la comunidad para la cual trabajan. El problema es que los costos suelen ser totalmente prohibitivas, no solo para las personas naturales, sino incluso para muchas de nuestras instituciones, lo que limita la difusión del conocimiento.

Por ello, desde inicios de este siglo, ha empezado un movimiento denominado de acceso abierto (open access) que promueve que los autores publiquen sus libros y artículos en instituciones que ofrezcan a sus usuarios el acceso abierto sin costo a su repositorio, para los usuarios que quieran visualizar su contenido.

Las ventajas para el autor que así procede son indirectas, pues ya no percibirá una retribución económica por derechos de autor, pero en cambio se hará conocido para muchísimas personas en todo el mundo (no solo en el ámbito en el que publica), lo que va a influir positivamente en su desarrollo profesional y académico.

6. Interoperabilidad.

Debido a las redes de comunicaciones, la información almacenada en cualquier lugar puede ser accedida desde otro lugar, siempre y cuando se tengan los permisos necesarios para hacerlo, y, sobre todo, la información esté almacenada bajo ciertos estándares universalmente reconocidos, es decir, que tanto el que depositó la información como el usuario que la accede hablen el mismo idioma computacional para entenderse. El protocolo que permite esto es el OAI PMH (Open Archives Initiative Protocol for Metadata Harvesting).

De esta manera basta que dos instituciones tengan un convenio de intercambio de accesos, para que los usuarios de ambas instituciones puedan acceder a todos los contenidos de sus respectivos repositorios.

Esta interoperabilidad también ha permitido la creación de repositorios nacionales, en los que la ley obliga, bajo determinadas condiciones, que los contenidos de los repositorios de cada institución sean también visibles en el repositorio de la institución del gobierno designada por la ley para tal fin. Este es el caso, en el Perú, del repositorio ALICIA, del CONCYTEC. A su vez los repositorios nacionales se consolidan en repositorios regionales, como es el caso de ALICIA frente a LA REFERENCIA.

7. Primer Repositorio de Tesis PUCP.

La PUCP siempre ha tratado de estar a la vanguardia en los temas tecnológicos. Por ello, desde fines del siglo pasado, ya promovió la presentación de tesis en soporte digital. A inicios de este siglo decidió la creación de un Repositorio de Tesis Digitales. La creación de este primer repositorio fue encargada a la Dirección de Informática, quienes desarrollaron el Repositorio sin tener en cuenta los estándares de biblioteca ni de interoperabilidad, por lo que el uso de este repositorio quedó supeditado a los miembros de la Comunidad PUCP y a aquellos que tuvieran

conocimiento de la existencia de este repositorio y accedieran por medio del URL www.tesis.pucp

Hacia mediados del 2010 este repositorio albergaba algo más de 400 tesis.

8. El Repositorio Institucional PUCP.

A fines del 2009 la PUCP encargó al suscrito analizar la problemática de los rankings universitarios. Como resultado de dicho análisis se determinó que la PUCP adolecía de falta de visibilidad en la web, y que tenía mucho que mostrar gracias a la riqueza de su acervo histórico y cultural, y a la calidad de los trabajos de investigación de sus docentes.

Para superar esta falta de visibilidad se propuso un plan de trabajo que se realizaría en tres etapas, el cual fue aprobado por el Rectorado (Consejo Universitario) a fines de junio del 2010. Este plan consistía en:

- a) Rediseñar el repositorio de tesis existentes, para dotarlo de los estándares de biblioteca y de interoperabilidad requeridos, así como crear los metadatos de cada tesis para hacerlas fácilmente accesibles.
- b) Generar un portal de revistas mediante el SW OJS para mostrar en un solo sitio todas las revistas editadas en la PUCP, tanto las correspondientes a versiones antiguas impresas, como las nuevas que ya se editaban en formato digital, pero que se alojaban en distintas páginas web.
- c) Crear un Repositorio Institucional donde se aloje el acervo histórico y cultural y la producción científico académica de la universidad.

9. Nuevo Repositorio de Tesis PUCP.

Entre julio del 2010 y abril del 2011 el personal de la Oficina de Automatización de la Biblioteca desarrolló un nuevo repositorio de tesis, con estándares de interoperabilidad y metadatos adecuados. En el interín siguió funcionando el anterior repositorio de tesis, pero éste ya no se siguió poblando con nuevas tesis.

Para este nuevo repositorio se recuperaron todas las tesis existentes en el repositorio anterior y se fueron añadiendo todas las nuevas tesis presentadas en las distintas facultades a partir de Julio del 2010. Pero lo más relevante es que, adicionalmente, se hizo un gran esfuerzo para recuperar las tesis de años anteriores que no estaban digitalizadas.

Debido a las regulaciones sobre derechos de autor, estas tesis antiguas no podían ser colocadas en el repositorio de tesis si no se contaba con una autorización escrita del autor, lo que nos obligó a ubicar uno por uno a los autores y solicitarles su autorización, habiéndose recuperado así alrededor de 1,300 tesis.

Este repositorio se ha seguido poblando desde entonces con tesis de bachillerato, licenciatura (título profesional), maestría y doctorado, conteniendo a la fecha más de 9,500 tesis.

Debemos resaltar que actualmente el procedimiento de titulación exige que el tesista autorice o deniegue desde un inicio la publicación de su tesis, como condición para que su tema pueda ser aprobado. En caso de denegación debe indicar las causas por las que lo deniega (por ejemplo patente con trámite en curso, implementación en la vida real del objeto de estudio, etc.), pero

no se le puede obligar a autorizar su publicación. Felizmente estos casos de denegación se presentan con muy poca frecuencia y en la mayoría de los casos no son definitivos, sino que tienen un periodo de embargo.

10. Portal de Revistas.

El siguiente paso fue desarrollar un portal de revistas, para lo cual se usó el SW OJS (Open Journal Systems).

Debemos resaltar en este caso que además de la tarea de programación hubo una ardua labor de “descubrir” todas las revistas institucionales que se editaban en la universidad y ubicar todos los números publicados de cada revista no digital, para proceder con su digitalización, tarea que resultó por demás complicada pues en la biblioteca no existían algunos números de determinadas revistas, por lo que se tuvo que recurrir a profesores y exalumnos para ubicar a alguien que si lo tuviera. Podemos decir con satisfacción que se logró el objetivo de ubicar las ediciones impresas de todos los números anteriores de las revistas vigentes en nuestro portal de revistas.

La presentación del portal de revistas se realizó en noviembre del 2012 y a la fecha cuenta con más de 40 revistas vigentes (es decir que se continúan editando).

11. Desarrollo del Repositorio Institucional (RI).

Luego del lanzamiento del Portal de Revistas nuestros esfuerzos se dirigieron a desarrollar y poblar nuestro repositorio institucional, el cual contempla incluir:

- Todas las Tesis ya registradas en el repositorio de tesis.
- Todos los artículos contenidos en las revistas del portal de revistas, incluyendo las pertenecientes a revistas que se fueran retirando del portal de revistas por haber perdido su vigencia, o a las revistas que al momento de lanzar el portal de revistas ya estaban descontinuadas pero cuyos contenidos siguen siendo de actualidad y de interés para la comunidad.
- Documentos de todo tipo (impresos, auditivos, gráficos, fílmicos) que componen el acervo histórico cultural de la PUCP.
- Producción científico académica de la comunidad PUCP, cuyo autor quisiera que se muestre en el repositorio.
- Material proveniente de proyectos de interés desarrollados por personas cercanas a la PUCP que tienen interés en mostrar sus resultados y con los cuales se establecieron convenios para su publicación en el RI de la PUCP.

El RI fue desarrollado usando el SW libre DSpace.

La información antes mencionada se organizó en comunidades, sub comunidades y colecciones según su fuente de origen, y, en algunos casos, según su contenido o tipo de soporte utilizado.

Una labor importante en la implementación del RI fue la digitalización del material existente en la PUCP, especialmente en el Archivo General y en la Biblioteca del Instituto Riva Agüero (IRA). Para facilitar dicho trabajo se instaló en el IRA un equipo (máquinas y personas) de Digitalización Fotográfica Cenital, el cual permite digitalizar fotográficamente y retocar

mediante el Photoshop documentos de diferentes tamaños y en diferentes estados de conservación. Este equipo se mantiene hasta la fecha y continua su labor de digitalización.

El lanzamiento del repositorio se realizó el 13 de noviembre del 2013, lo cual no significa que se haya concluido el trabajo, pues la labor de recuperación del acervo histórico cultural de la Universidad excede largamente nuestra capacidad actual y no disponemos de mayores recursos para hacerlo más rápido.

En el RI de la PUCP se han acogido, mediante convenios, dos importantes proyectos:

- PESSCA (Project on the Engraved Sources of Spanish Colonial Art), de nuestro exalumno e investigador Almerindo Ojeda, cuyo objetivo es establecer la correspondencia entre las pinturas producidas durante la colonia (especialmente en Perú y Ecuador) y las pinturas existentes en Europa, en las que supuestamente se inspiraron los autores coloniales. A la fecha este proyecto muestra más de 4,000 de estos emparejamientos.
- Glosas Croniquenses. Proyecto desarrollado por la Dra. Lydia Fossa, que se inició en la Universidad de Arizona (Phoenix) y se continuó en la PUCP, por ser de interés para el estudio de las lenguas andinas. En este proyecto se analizan las palabras de distintos lenguajes autóctonos (quechua, aimara, taino, entre otros) que utilizaron en sus relatos los distintos cronistas de los primeros tiempos de la colonia, desde la conquista hasta finales del siglo XVI.

Se inició con las palabras usadas por 7 cronistas y actualmente se ha llegado a estudiar a 14 cronistas.

12. Recursos utilizados.

Para el desarrollo de este repositorio se formaron dos equipos de trabajo:

Uno que tiene que ver con los contenidos, es decir, ubicar la información susceptible de formar parte del repositorio y establecer prioridades para su inclusión y otro que tiene que ver con el desarrollo del SW requerido y la correcta asignación de metadatos para registrar los diversos documentos en el repositorio.

El primer equipo pertenece a la Asesoría Técnica del Rectorado (ATR) que yo lidero y el segundo al Sistema de Bibliotecas. Para coordinar el trabajo de ambos equipos existe un comité de coordinación que se reúne mensualmente (antes, hasta hace un año, lo hacía quincenalmente).

Por razones prácticas el equipo de digitalización fotográfica que trabaja con el material del IRA pertenece a la ATR, a mi cargo mientras que el Sistema de Bibliotecas cuenta con su propio equipo de digitalización fotográfica.

En el equipo de la ATR participan además del suscrito un Jefe de Contenidos, 1 experto en TIC's, 1 técnico en digitalización fotográfica y dos practicantes.

En el equipo del Sistema de Bibliotecas participan, a tiempo parcial, compartiendo su labor con las labores propias de la Biblioteca, el jefe de automatización, dos analistas programadores (expertos en DSpace, OJS y otros), 1 responsable de las revistas y 1 responsable de catalogación y registro de contenidos. Debemos resaltar que también se cuenta con sendos responsables de catalogación y registro en cada comunidad para que ingrese al repositorio los documentos

propios de su comunidad, los cuales son previamente entrenados para ello por el responsable del Sistema de Bibliotecas.

En cuanto al HW requerido, actualmente se utilizan tres servidores contratados en la nube con Amazon Web Services (como parte de un contrato global de la PUCP con AWS): uno para los programas, otro para los datos y un tercero para alojar el proyecto PESSCA, el cual actualmente está alojado en paralelo en un servidor situado en Davis (California).

13. Conclusiones y Recomendaciones.

- Los Repositorios en general pueden contener cualquier tipo de información. Los Repositorios Institucionales en cambio deben contener únicamente la información que mejor represente a la institución. No deben reducirse a mostrar las tesis, aunque esta es una buena manera de iniciarlos, como ha sucedido en la PUCP.
- Los RI, deben ser de acceso abierto para las consultas, es decir deben poder ser consultados desde cualquier lugar, por cualquier persona, sin ninguna restricción ni necesidad de identificación previa, pero en cambio deben ser cerrados en cuanto al registro de documentos, para mantener su carácter institucional. Es decir, la nueva información solo puede ser registrada por el personal autorizado y debidamente capacitado de cada unidad.
- Se deben preservar los derechos de autor de los autores que aceptan que su información se registre en el RI. Esto significa comprometer a los usuarios a no utilizar con fines comerciales el material del RI y hacer siempre referencia al autor y al RI de la PUCP cuando utilicen (sin fines comerciales) la información contenida en él.
- La labor de poblamiento del RI es una labor continua, tanto por la incorporación de nuevos materiales que se van generando, como por la inclusión de los materiales que componen el acervo histórico cultural de la institución conforme se vayan digitalizando.
- Tan importante como poblar el repositorio es asegurar el que sea consultado, es decir que la comunidad nacional e internacional lo adopte como una valiosa fuente de información para sus investigaciones.

Para ello es necesario difundir su uso con presentaciones globales, como ésta, o desarrolladas con fines específicos, para mostrar algunos de sus contenidos, y llevar una adecuada estadística de su uso.

Por ejemplo el RI se inauguró en el 2013 con 13 comunidades y 10,000 documentos y a la fecha tiene más de 30,000 documentos, pertenecientes a 32 comunidades, siendo visitado mensualmente por unas 15,000 personas, resultado de una expansión anual en número de visitas cercana al 50%.

- Una última recomendación es revisar constantemente la calidad e idoneidad de sus contenidos y asegurarse el correcto funcionamiento de todo lo que ofrece el RI, pues un usuario insatisfecho puede resultar en una pérdida de otros usuarios de su entorno y causar un desprestigio para la institución.

Lima, 24/10/2018