

El buscador de recursos de la UNSAM. Apuntes de una experiencia

Higa, Diego

Pastrana, Luis

Uviña, Ramiro

Introducción: la UNSAM

La Universidad Nacional de General San Martín (UNSAM) está ubicada geográficamente en el Partido de General San Martín, Provincia de Buenos Aires, y fue creada en el año 1992. Siendo pública y gratuita, ofrece un abanico de carreras de grado y posgrado, en diversas disciplinas del hacer y el conocimiento.

Como ya mencionamos su sede está ubicada al noroeste de la Ciudad Autónoma de Buenos Aires. El campus Miguelete está construido sobre una antigua playa ferroviaria de más de dos hectáreas, constituyéndose como uno de los principales atractivos del territorio bonaerense en términos de planeamiento arquitectónico y conservación patrimonial¹.

La Biblioteca Central

Comienza a funcionar a fines del año 2010, y cuenta con un edificio de 1.000 m² distribuidos en tres plantas, con capacidad para alojar hasta 70.000 volúmenes en estantería abierta y con trescientos puestos de lectura².

Su colección inicial integró tres bibliotecas pertenecientes a la Escuela de Ciencia y Tecnología, la de la Escuela de Humanidades y la del Centro de Estudios Latinoamericanos³. A partir de allí su patrimonio bibliográfico se incrementó progresivamente (año a año) a través de compras, donaciones de figuras de las ciencias y la cultura y el acceso a bases nacionales e internacionales.

El catálogo cooperativo de la UNSAM

La Biblioteca Central de la Universidad Nacional de San Martín (BCUNSAM) coordina un catálogo cooperativo conformado por las colecciones de la: Biblioteca Central (BC), Biblioteca del Centro TAREA, Biblioteca de Posgrado - EPYG -

¹ <https://www.unsam.edu.ar/institucional/institucional.asp>

² https://www.unsam.edu.ar/biblioteca_central/sobre_biblioteca1.asp

³ <https://bibliotecadigital.cin.edu.ar/bitstream/123456789/2215/5/Art%C3%ADculo%20desarrollado>

Escuela IDAES, Biblioteca de la Escuela de Economía y Negocios y del Instituto de Ciencias de la Rehabilitación y el Movimiento⁴. El mismo contiene más de 68.000 registros bibliográficos, entre los cuales se encuentran diferentes materiales, incluyendo algunos electrónicos, de todas las bibliotecas que participan en el mismo.

Este catálogo utiliza el software Koha, sistema integrado de gestión bibliotecaria libre y gratuito. Utiliza Perl como lenguaje de programación, una base de datos relacional (MySQL o MariaDB) y en el empaquetado se incluye Zebra como motor de búsqueda, pudiendo utilizarse Elasticsearch para los mismos fines⁵.

Siendo un sistema integrado de gestión bibliotecaria incluye los módulos de catalogación de materiales y de existencias, de autoridades, gestión de usuarios y de circulación, entre otros. Su objetivo principal es gestionar los recursos de una biblioteca tradicional en forma centralizada, integrando información de diferentes tipos en un único software.

Entre sus características destacadas pueden mencionarse la utilización de MaRC21, estándares de interoperabilidad OAI-PMH, SRU y la posibilidad de acceder a reportes vía JSON. Además ofrece una API REST en desarrollo⁶.

El Repositorio Institucional de la UNSAM

Se inició en el 2010. Luego, en el año 2014, por resolución CS 325/14 se oficializa su existencia.

Desde sus comienzos se gestionó utilizando el software Greenstone 2.82 como herramienta de gestión. En los primeros momentos se contaba con un acceso FTP al servidor del mismo para poder acceder a las carpetas y subir o bajar archivos desde la Biblioteca Central. Su primera colección fue exclusivamente de tesis.

La carga era eminentemente centralizada, dadas las características del software de gestión del mismo.

En 2019 se implementa y migra del Greenstone 2.82 al DSpace 6.3, en esta oportunidad desde la Biblioteca Central se solicita acceso SSH con usuario perteneciente al grupo *sudoer*, con lo cual se instala el DSpace 6.3, se configuran las comunidades, subcomunidades y colecciones y luego se migra de un software a

⁴ <http://koha.unsam.edu.ar/>

⁵ <https://koha-community.org/about/>

⁶ <https://koha-community.org/manual/23.05/es/html/webservices.html#>

otro desde la Biblioteca Central. Esta nueva implementación permitió una participación activa desde las Bibliotecas de la UNSAM y las sedes de las Escuelas de la Universidad que gestionan las tesis en texto completo para que pudieran cargarlas (ya se habían incluido libros electrónicos y artículos de investigación además de las tesis y trabajos finales).

En junio de 2021 el Sistema Nacional de Repositorios Digitales (SNRD) adhiere al Repositorio Institucional (RI) de la UNSAM en forma provisoria, estando en la actualidad en la fase final de la adhesión total. Por otro lado, también se está trabajando en la política institucional de acceso abierto.

A fines de ese mismo año se opta por incorporar al RI a los investigadores y sus proyectos, así como las financiaciones relativas a los mismos de un modo más visible. Para ello, se instaló una instancia de DSpace-CRIS y se migró de la versión DSpace 6.3 a este nuevo software, a partir de lo cual se cargan algunos registros de investigadores. Estando en la actualidad todavía incorporando artículos de los mismos.

El Portal de Revistas Académicas de la UNSAM

En febrero de 2016, el Portal estuvo en línea por primera vez. Son siete las revistas que actualmente la integran: artículos, dossiers temáticos, avances de investigación, reseñas y entrevistas forman el contenido de las revistas. Estas son parte de las más de 33 mil que utilizan el software libre Open Journal System (OJS), el más usado y aceptado para gestionar este tipo de publicaciones en todo el mundo. En sus comienzos, el Portal consistía de una serie de páginas web y documentos PDF subidos utilizando el lenguaje de programación ASP, y dependía de la Secretaría de Investigación de la UNSAM.

En el año 2017, se implementa OJS 2 como software de gestión de las revistas académicas, dando un salto de calidad en lo relativo a la subida de números y gestión de usuarios, así como en la descentralización de los procesos.

En el año 2020, el proyecto de gestión de revistas académicas de la Universidad se traspa a la Biblioteca Central, llevando consigo una serie de problemáticas en lo descriptivo y en el software en sí mismo. Por ello, se solicita gestionar el software con acceso *root* al servidor, en el cual se alojan las publicaciones. La primera acción que se toma es actualizar el versionado del OJS. Para ello, se instaló una nueva versión de OJS 3 y se migró la base de datos al mismo, corriendo las

actualizaciones de SQL correspondientes. Una vez realizada esta actualización se fueron resolviendo los errores, que anteriormente no habían podido solucionarse de manera paulatina. A partir de esta implementación exitosa se generó una nueva área dentro de la Biblioteca Central, denominada Desarrollos Informáticos; la cual está constituida por tres bibliotecarios que se dedican a tiempo completo a las cuestiones tecnológicas relativas a las Bibliotecas de la UNSAM y también con otras externas con las que se realizan convenios.

Qué es el Vufind

El Vufind es una capa de descubrimiento y un motor de búsqueda⁷. Software gratuito de licencia de código abierto, mayormente utiliza PHP como lenguaje de programación, una base de datos relacional (recomiendan utilizar MySQL o MariaDb, existiendo la posibilidad de utilizar PostgreSQL) y requiere de JDK para el motor de búsqueda, que es Solr.

Sus objetivos son variados, pero todos apuntan a ofrecer un modo de consulta de las colecciones bibliotecarias y afines que tome los datos de otro sistema para ofrecerles una interfaz de usuario independiente de los mismos⁸. La denominación “bibliotecarias y afines” es porque, como se viene desarrollando en esta ponencia, en la actualidad mencionar que las bibliotecas poseen materiales que se describen utilizando las reglas de catalogación (sean las angloamericanas, las de acceso y descripción o cualquier otro modo que refleja la biblioteca tradicional) es perder de vista que las bibliotecas hoy poseen un patrimonio mucho más extenso, el cual es fruto de la incorporación y reconocimiento de las unidades de información a la organización mayor en la cual está inserta (en nuestro caso, en la universidad). En este sentido, Vufind permite ofrecer a través de una boca de consulta única los recursos de los sistemas integrados de gestión bibliotecaria, el sistema de repositorio institucional, de revistas o bases de datos comerciales, así como muchos otros.

⁷ <https://vufind.org/docs/VufindBrochure2.pdf>

⁸ <https://github.com/vufind-org/learning-vufind-book/releases/download/v1.1.2/LearningVuFind.pdf>

Mención aparte merece el modo en que Vufind toma los datos de otros sistemas. Si bien, en general cosecha los datos a través del protocolo OAI-PMH también puede importar registros (mayormente de MaRC21 o MARCXML, en formato .mrc) a través de la copia de archivos a su servidor y su posterior indizado en Solr.

El Vufind en la Biblioteca Central: el buscador de recursos de la UNSAM

Desde la dirección de la Biblioteca Central de la UNSAM la implementación de un software de descubrimiento era un proyecto en carpeta desde hacía algunos años, pudo concretarse recién cuando surgió una nueva área en la BC.

Para el Buscador de Recursos de la UNSAM se incluyeron únicamente las colecciones propias de la biblioteca, excluyendo las bases de datos comerciales suscritas por un tiempo definido. De este modo, el catálogo de la biblioteca, el repositorio institucional y el portal de revistas académicas son las plataformas que integran el Buscador de Recursos de la UNSAM.

Una vez definido el alcance del proyecto, se revisó la configuración de cada sistema por separado. Como primera medida, hubo que verificar si el protocolo OAI-PMH estaba habilitado y funcionando correctamente; luego hubo que definir y revisar cómo estaban dispuestos los sets o conjuntos de información. Por último, se debió planificar cada cuanto se realizaría la actualización de los datos cosechados.

Por supuesto, una vez definido el alcance del proyecto y destinados los trabajadores a tal fin (el área de desarrollos informáticos de la BC con asistencia de todos los sectores de la biblioteca en diferentes roles, más otros sectores externos a la BC), para comenzar el proyecto, luego de realizar la solicitud, desde el área de Sistemas se otorgó acceso a un servidor virtualizado específico, un Debian 11 con características mínimas de server: 2 vCPU, 8 GB de RAM y 25 GB de almacenamiento, en línea con la documentación oficial de Vufind⁹. Toda la instalación y configuración inicial del sistema corrió por cuenta del sector Desarrollos Informáticos de la BC. La versión estable del Vufind al momento de la implementación fue la 8.1 y esta es la versión que funciona hasta la actualidad.

⁹ <https://vufind.org/wiki/installation:requirements> y <https://vufind.org/wiki/administration:performance>

Analizando la integración de Vufind con Koha

El primer software con el cual se trabajó la integración de Vufind fue el Koha, que ya tenía el servidor OAI-PMH habilitado y funcionando correctamente (expone casi todos sus registros en un gran set genérico). De todas formas, hubo que enfrentarse a determinados dilemas:

-Si bien Vufind toma algunos datos de la base de datos de Koha directamente, por una cuestión de política de seguridad de la UNSAM no era posible sincronizar la información de ambos servidores (el de Koha con el de Vufind), por lo cual aquellos datos que no fueran a ser cosechados a través del protocolo OAI-PMH debían ser exportados desde Koha al servidor de Vufind a través de un dump de la base de datos.

-Las imágenes de las portadas de los libros eran demasiado pesadas para ser exportadas al servidor de Vufind utilizando este método, ya que como fue mencionado el servidor posee mucha capacidad de procesamiento pero no tanta de almacenamiento¹⁰.

-Algunas políticas de visualización de campos de ítems de Koha (estados, colecciones, etc.), se despliegan de un modo mientras que en Koha se almacenan de otra forma.

-Al ser el software que tiene gestión de usuarios más extendida, se lo toma como base para la autenticación de los mismos.

Analizando la integración de Vufind con OJS

El segundo software que se consideró para integrar al Vufind fue el OJS, correspondiente al Portal de Revistas Académicas. Al igual que se realizó con Koha, se analizó cómo el OJS desplegaba los datos.

¹⁰Todavía no estaba implementada la funcionalidad de tomar las imágenes a través de una consulta web, la cual se entiende que es posible a partir de la versión 9.

El primer aspecto que se visibiliza es que cada revista tiene su propia URL para desplegar los registros en OAI-PMH¹¹. El segundo aspecto que se destaca es que los sets como tales se relacionan con los apartados de cada publicación¹²: así como presenta un conjunto con todos los registros de la publicación.

Por otro lado, una cuestión también evidente es la normalización utilizando vocabularios OpenAIRE de los campos tipo y MIME. Para los idiomas utiliza la norma ISO 639-2, al igual que Koha.

El servidor OAI-PMH de este software no despliega información vinculada a qué apartado de la revista pertenece cada registro, y solamente despliega artículos. Esto es relevante ya que gracias al campo *source* existe la información contextual de cada registro (a qué título y número pertenece), pero no a qué sección.

Analizando la integración de Vufind con DSpace-CRIS

El último software que se analizó fue el DSpace-CRIS, con el cual ya se había trabajado el despliegue de metadatos en OAI-PMH para que la UNSAM pueda incorporarse al SNRD. Es por este motivo, que se estableció *a priori* que sería el más sencillo de integrar a Vufind. Incluso, por este mismo motivo se utilizó el mismo set OAI-PMH denominado SNRD para tomar los registros, que son trabajados y evaluados continuamente.

DSpace también utiliza vocabularios OpenAIRE de los campos tipo y MIME para el formato, al que le agrega el vocabulario SNRD para la tipología en español. Para los idiomas utiliza la norma ISO 639-2, al igual que los otros dos softwares.

El servidor OAI-PMH de este software no despliega información sobre la jerarquía a la que pertenece cada registro, solamente lo relacionado a los ítems. Esto es relevante ya que utilizando este método se pierde la relación de pertenencia de la comunidad, subcomunidad y colección con cada registro.

¹¹ <https://docs.pkp.sfu.ca/admin-guide/en/data-import-and-export#oai>

¹² <https://revistasacademicas.unsam.edu.ar/index.php/etnocontemp/oai?verb=ListSets>

Evaluando la cosecha de Vufind

Cuando se integró DSpace al Vufind ya se habían realizado pruebas de cosecha de Koha y OJS, por lo cual ya era posible revisar los resultados completos.

Con todos los registros cosechados, se revisaron los resultados donde finalmente para todos los softwares hubo que realizar ajustes.

Cosechado con Vufind de Koha

Contra todo pronóstico, quizás uno de los más complejos fue Koha, dado el dinamismo que presenta la disponibilidad de los materiales registrados en el mismo. Para este caso puntual y como no era posible conectar el Vufind con la base de datos de Koha en forma directa, se creó un script que exportaba determinadas tablas de Koha (se comenzó exportando *biblio*, *branches*, *issues* e *items* para luego agregar información de reservas y de usuarios) a un archivo. Luego se programó una tarea que de manera diaria copia este archivo en el servidor de Vufind para luego importar esta mini base de datos (con únicamente las tablas exportadas) en el gestor de bases de datos de Vufind (MariaDB). Por otro lado, se exportó toda la base de datos bibliográfica junto con sus ítems a un archivo .mrc para luego procesarlo con el SolrMarc de Vufind. Cabe destacar que Koha no está instalado en el servidor del Vufind, así como tampoco el Zebra o el Elasticsearch, únicamente existe una base de datos con algunas tablas de Koha en forma redundante.

Además, hubo que seleccionar entre los dos esquemas de metadatos que Koha permite, Dublin Core y MARCXML. Por obvias razones, se definió utilizar MARCXML.

Al haber todavía cuestiones puntuales que no terminaban de convencer y, aprovechando que se realizaba la copia de bases de datos de Koha al gestor de MySQL de Vufind, se armó una consulta de actualización de registros de ítem para que muestre los resultados de forma específica y personalizada, sobre todo en lo

relativo a algunos materiales que no se prestan a domicilio con alguna característica distintiva.

Uno de los puntos que se tiene en cuenta en la implementación realizada es que la disponibilidad de los ítems físicos no se refleja en Vufind en tiempo real, siendo lo único que se podría considerar para mejorar en el caso de la integración de Koha con Vufind, así como sucede con las claves de los usuarios en Koha.

Cosechado con Vufind de OJS

En el caso de OJS, lo más problemático fue que al ser un software multilingüe (se cargaron los datos en todos los idiomas que presentan los artículos), el software en la versión utilizada presentaba para algunos campos los resultados de OAI-PMH en inglés preferentemente (el *abstract* y el título era lo que más generaba inquietud) y obviaba o dejaba en segundo término el español. En este caso puntual, se revisó toda la configuración, además de foros en la web, sin dar con una respuesta. Finalmente se debió modificar el código de un archivo .php para que tome la preferencia de localización predeterminada al momento de desplegar los datos.

Posteriormente, se observó que tanto la tipología de los artículos como el idioma de los mismos aparecían utilizando diferentes vocabularios controlados. Esto “ensuciaba” un poco los filtros para el usuario final (el SolrMarc los reformatea para que los idiomas aparezcan en forma simple y desarrollada -español, inglés, italiano, etc.-). Esto mismo sucedería para DSpace.

Cosechado con Vufind de DSpace-CRIS

Por otro lado, en lo relativo a la cosecha de los registros del repositorio institucional se tuvieron algunas problemáticas idénticas que con OJS, y que era el momento de enfrentar: los vocabularios controlados son una maravilla que nunca van a ser ponderados como corresponde, pero al momento de desplegar la información a los usuarios se consideraba de interés que los datos aparecieran en la forma más coloquial posible. Es decir, que en lugar de “spa” en el filtro de idiomas apareciera

“Español” y para el caso de la tipología documental en lugar de “info:eu-repo...book” apareciera, por ejemplo, “Libro”.

Afinando la cosecha de Vufind

Fue así que se comenzó a revisar toda la configuración del Vufind en búsqueda de tablas de equivalencias, reformato o lo que se tuviera al alcance. Como no había conocimiento del Solr al respecto, se investigó también ese aspecto, sin descartar ninguna opción.

De esta forma, se concluyó que existían varias alternativas:

- a) Modificar cómo desplegaba los datos el OAI-PMH de DSpace u OJS.
- b) Intervenir los datos al momento de ser cosechados del software específico (DSpace u OJS) al Vufind, es decir, modificar el script de cosecha de registros en Vufind.
- c) Intervenir el índice de Vufind una vez que los registros ya fueron cosechados (el solr).
- d) Explorar las herramientas de interoperabilidad fuera del OAI-PMH.

Finalmente, se optó por un mix de las alternativas. En algunos casos, como el despliegue en otros idiomas de OJS se intervino el código de dicho software, o aquellas consultas de ítems para el Koha; se realizó esto cuando se identificó que no había ninguna otra forma de obtener resultados satisfactorios. En los casos que parecían simples y hubo conformidad con los resultados, se elaboraron tablas de equivalencias y reconversión de datos en el momento de la cosecha del software específico al Vufind. Y en aquellos casos que el punto “b”, no alcanzó, o no se supo configurar correctamente, se intervino el índice de Vufind a través de scripts que se dejaron programados para que se ejecuten una vez que se actualizan los datos tanto de DSpace como de OJS.

Si bien esto último fue quizás el tema que más tiempo demoró en llevarse a cabo, se aprendió mucho en el proceso, y siempre se tuvo como principal activo que

desde el sector todos tienen acceso a los softwares mencionados y sus datos en forma directa, a través del acceso a los servidores con sus bases de datos.

A modo de conclusión: un trabajo en equipo

Por supuesto, que por cada nueva cosecha o solución implementada desde el área de Desarrollos Informáticos con respecto a Vufind todo el resto de los trabajadores y trabajadoras de la BCUNSAM aportó su parte: haciendo de testers, observando errores y discrepancias en los datos, así como consultando acerca de opciones o definiendo qué había que retirar de la vista pública.

Además, ya en la etapa de finalización de la implementación de Vufind en lo que a datos se refiere, se hizo el contacto con el área de diseño de la UNSAM, quienes de forma rápida le dieron forma al nuevo Buscador, tal como fue denominado en base a su nombre de dominio, que desde el área de Sistemas fue brindado y posibilitado utilizar una vez que se salió a producción. Esto es lo que resulta quizás más interesante del Descubridor implementado: además de lo que se aprendió en el camino, de cómo los integrantes de la flamante área de Desarrollos Informáticos estrenaron y se probaron a ellos mismos integrando todos los recursos de datos de la biblioteca, se rescata que el buscador hoy es el reflejo de todo lo que se realiza en la UNSAM, por la totalidad de sus trabajadores y trabajadoras. Desde el artículo o referato de una publicación académica para los profesores e investigadores, pasando por las tesis de los estudiantes de grado o posgrado, llegando a toda la labor que se realiza en la Biblioteca y sus instituciones adheridas en cuanto a la descripción de recursos. Por supuesto, también participaron los trabajadores y trabajadoras de otras áreas como Sistemas y Diseño, que desempeñaron roles no menores.

Es por todo esto, el orgullo de haber realizado esta tarea y poder comunicarla. Esta experiencia habla también de lo importante que es contar con un equipo de personas comprometidas y dispuestas a aportar su parte de conocimiento al conjunto. A su vez, para que todo esto sea posible y la comunidad educativa sea la beneficiada (sea de forma directa o indirecta, ya que el descubridor posicionó nuestros recursos mejor aún en los algoritmos de ranqueo de los sistemas de

recuperación de la información más utilizados en la web) es imprescindible sostener y mejorar la inversión en estas instituciones de educación, difusión y transferencia de conocimientos, siendo este último punto el que se intenta honrar hoy con esta ponencia.