

GPLib: El encuentro de lo tangible y lo intangible en el Software Libre

El proyecto de creación de un sistema libre de
Gestión y Repositorio de Objetos de Conocimiento.

Autor: Pablo Javier Etcheverry

Institución: Universidad de Buenos Aires – Facultad de Filosofía y Letras

Índice de contenido

1. Introducción
2. Situación - Problema
3. Solución
4. Innovación e Inédito
5. Beneficiarios
6. Relevancia para el Interés Público
7. Viabilidad Técnica, Financiera y Política Organizacional
8. Facilidad de Reproducción
9. Ambiente de Hardware y Software

Introducción

En los últimos 20 años, han nacido y desaparecido una importante cantidad de software relacionado con las Bibliotecas, por un lado, y con la gestión de documentos digitales, por el otro. Algunos de ellos han perdurado, tanto dentro del paradigma del software privativo como en el Software Libre (SL). Sin embargo, hasta el momento no existía un sistema que combinara ambas funcionalidades.

Esta separación no ha sido casual sino perfectamente adecuada. En efecto, los comúnmente conocidos como *Sistemas bibliotecarios* (SB) no han tenido sino un lugar muy específico, circunscripto a la gestión de acervos físicos: libros, revistas, diarios, mapas y algunos objetos más, pero siempre en el dominio de lo concreto, siempre pensando el objeto cultural como físico, único y -sobre todo- prestable. Estos supuestos, sobre los que el diseño de todos los SB descansó, estallan con la introducción del formato digital.

El mundo digital, la digitalización de estos mismos elementos (libros, revistas, diarios, mapas) conlleva el cuestionamiento de lo tangible y estos nuevos objetos intangibles son por su propia naturaleza múltiples y expandibles; pueden ser copiarse, replicarse, compartirse sin limitaciones más que las técnicas y legales.

Esta sorpresa no llegó de la misma forma a las comunidades que conocían o creaban *Software de Repositorios* (SR). Ellas conocían desde su propia concepción los objetos digitales. Los SR se especializan en tomar objetos digitales, preprocesarlos, agregarles metadatos y almacenarlos. Una vez almacenados pueden realizar búsquedas sobre ellos, recuperarlos, fusionarlos o entregarlos al usuario en un formato diferente del que tenían al ingresar al repositorio (ej: puede obtenerse un PDF aún cuando el archivo original fuera un TXT).

Así como los SR entendían la naturaleza digital, no comprendían que existieran “activos bibliográficos”, “incunables”, o simplemente libros. Las masas de papel acumuladas y sobre todo administradas diariamente por personas que transitan en pasillos de estanterías no le interesaban a las comunidades de entusiastas repositoristas. Ellos se concentran en lograr la más alta eficiencia en almacenar y recuperar objetos digitales, pero cualquier proceso vinculado con lo físico, o con la administración de estos recursos es ignorado.

Nuevamente nos encontramos con una actitud adecuada. Si se manejan bienes intangibles donde su copia y reproducción no afecta el patrimonio de la institución que los administra, entonces ninguna de las

preocupaciones de los SB (el tiempo que un ítem lleva prestado, las sanciones al deudor, el costo de reponer una copia) puede estar presente.

Esta desconexión que también se presenta en otros ámbitos (ERPs y CRMs, por ejemplo) podría haber seguido existiendo si no fuera por la aparición de una necesidad concreta. En este caso, una institución estatal de educación decidió que la mejor forma de administrar y poner a disposición de un público más amplio su conocimiento acumulado era crear un sistema que permitiera cumplir con ambas funciones. No sólo esto: decidió que este sistema debía ser creado bajo el paradigma de Software Libre, para que otras instituciones pudieran aprovechar la inversión que los ciudadanos harían en su creación.

Situación – Problema

En el contexto de la Facultad de Filosofía y Letras de la Universidad de Buenos Aires la Biblioteca Central (BC) fue históricamente un lugar relegado en cuanto a la incorporación de tecnología. No profundizaremos aquí en las razones para ello más allá de mencionar que la escasez de fondos colocó desde los años 90 a esta área en el fondo de la lista de prioridades de asignación de recursos técnicos.

Al momento de comenzar el proyecto que motiva esta presentación, dicha área contaba con cuatro sistemas de gestión de datos:

- 1) Sistema de gestión de préstamos. Desarrollado por un programador independiente a mediados de los 90 , este sistema se utilizaba para registrar el préstamo de libros. No se contaba con el código fuente ni ninguna documentación formal. La base de datos era propia y la carga de usuarios, manual.
- 2) Sistema de información de libros. Construido utilizando el sistema CDS/ISIS, este catálogo registraba el patrimonio bibliográfico de la Biblioteca Central (BC). La base de datos estaba completamente desconectada del sistema de préstamos y la cantidad de elementos registrados en ambos sistemas difería en forma alarmante.
- 3) Catálogo en línea. En base al sistema ISIS, se construyó en el año 2009 un OPAC (Online Public Access Catalog) que se colocó en línea a disposición de cualquier usuario que deseara conocer las existencias de libros. Internamente, se dispusieron 4 PC para que los usuarios de la BC pudieran conectarse al catálogo y realizar búsquedas. Dicho catálogo no actúa en línea sino que se actualiza periódicamente con los datos obtenidos de ISIS.
- 4) Fichas de cartón. Son la forma más tradicional de búsqueda y no sólo se conserva sino que es utilizado a diario por el personal de atención al público.

Vemos que es evidente la superposición de funciones entre los cuatro sistemas. No sólo eso, sino que no hay conexión entre ninguno de ellos, con excepción de una relación de volcado del (2) al (3). Entonces, además de superposición de funciones hay una grave falencia respecto de la integridad de datos.

La necesidad entonces se define en forma general como la integración de dos grandes funciones, gestión de procesos y repositorio digital. El resultado deberá respetar en forma general los siguientes puntos:

- Integración de los datos y sistemas existentes en un único sistema nuevo.
- Almacenaje de las tesis de doctorado y de maestría, los proyectos de investigación, obras del tesoro y cualquier otra producción que conforme la memoria colectiva de la Facultad.
- Solución para la búsqueda de material por cualquier campo.
- Búsquedas de texto completo en los materiales digitales cargados, además de la publicación del material original, tal como fuera producido.
- Autenticación de usuarios en forma centralizada.
- Sistema de gestión de roles, con permisos diferenciados, donde cada usuario pueda recibir uno o más roles.
- Circuito de adquisiciones: compra, canje, donación u otro ingreso de ítems, con responsabilidades repartidas entre los diferentes actores del circuito (compras, dirección, catalogación, etc.).
- Todo el sistema debe respetar MARC y debe ser adaptable a otras normas.
- Comunicación via OAI, tanto para descarga de datos de catálogos como Worldcat como para entrega de datos a otros consumidores.
- Circulación y préstamo, reserva, control de tiempos, interfaces con sistemas de gestión académica.
- Calendario y generación de eventos.
- Interfaz con el Entorno Virtual de Aprendizaje Moodle.

Solución

GPLib es un sistema desarrollado a partir de un contrato generado a través de una licitación pública. La empresa contratada deberá entregar todo el sistema funcionando junto con su documentación completa. Sin embargo, y como se detallará en el punto de Innovación, la forma de trabajo no será monolítica, ya que para la implementación se ha optado por metodologías ágiles en lugar de desarrollo en modalidad *waterfall*. Además, se ha garantizado un espacio para la participación de todos los usuarios finales e intermedios.

Teniendo en cuenta ese marco, el desarrollo se ha planteado en ocho grandes fases:

- **Fase 1: Puesta en marcha.**
 - Objetivo: definición de las bases del sistema (core), las bases de datos, las fuentes de datos de terceros y la autenticación.
 - Entregable: Administrador de base de datos para comprobar la estructura creada ,código fuente del framework implementado y modificado,usuarios y grupos de pruebas,consultas simples para comprobar conexión de base de datos de terceros (Moodle y Greenstone), administrador web para la configuración y administración del sistema,módulo de login con ACL.
- **Fase 2 : Motor de Workflow y Sistema de Carga de existencias.**
 - Objetivo: Desarrollar un motor de workflow flexible y configurable capaz de respetar todos los estándares vigentes para adaptarlo al flujo de información.
 - Entregable:Módulo básico para la carga de existencias,módulo completo para el diseño e implementación de workflows.
- **Fase 3: Adquisiciones.**
 - Objetivo: Desarrollar los subsistemas de carga de existencias utilizando los workflows desarrollados específicamente para cada tipo de adquisición:compra,canje y donación.
 - Entregable: Módulo completo de adquisición y control de existencias .
- **Fase 4: Catalogación.**
 - Objetivo: Desarrollar e implementar un módulo de catalogación respetando los estándares MARC21, Z39.50 y SRU/SRW.
 - Entregable: Módulo completo del sistema de catalogación.
- **Fase 5: Búsquedas.**
 - Objetivo: desarrollo de un sistema de búsquedas avanzadas completo,que permita su utilización

en cualquier parte del sistema.

- Entregable: Módulo completo de búsquedas. Incluye Biblioteca Digital + OPAC y la implementación de una interfaz web pública y administración de usuarios (Moodle, Greenstone y LDAP).

- **Fase 6: Circulación y calendario.**

- Objetivo: Desarrollar dos módulos del sistema :un módulo de Circulación y un módulo Calendario, que interactúe con los módulos anteriores y permita manejar la calendarización y control de la circulación de existencias.
- Entregable: Módulo completo de circulación, módulo completo de calendario.

- **Fase 7: Reportes.**

- Objetivo: Desarrollar un módulo de reportes que pueda ser utilizado en todas las funcionalidades y módulos que requieran el uso de informes y reportes.
- Entregable: Módulo completo de reportes e informes.

- **Fase 8: Puesta en producción**

- Objetivo: Poner en producción el sistema completo parametrizado, configurado y con los registros migrados de los sistemas anteriores.
- Entregable: Sistema completo funcionando en producción, código fuente de todo el proyecto, informe final sobre el desarrollo completo y toda la documentación generada.

Innovación e Inédito

GPLib se presenta como un producto inédito en los siguientes aspectos:

- Es el primero en integrar dos necesidades como lo son la gestión de las operaciones en bibliotecas y los repositorios de conocimiento.
- Es el primer software que, generado desde una institución pública, es concebido desde sus inicios para ser liberado para su uso por parte de cualquier institución que lo requiera, considerando que el hacerlo no genera perjuicio para la institución que libera sino todo lo contrario, suma mano de obra en los procesos de soporte y *debugging*, por citar sólo dos ejemplos. Ésta se convierte en la innovación más importante en lo que respecta al ámbito público. En este punto cabe reconocer dos antecedentes. En primer lugar el del framework SIU-Toba, que gracias a la resolución del Ministerio de Educación, Ciencia y Tecnología número 823 del 25/6/2007 fue licenciado con el permiso de su libre utilización. Aún cuando fue un gran avance en su ámbito, debemos destacar que el hecho de no explicitar la licencia GPL (aún cuando los términos se superponen casi por completo) impide su integración con otros sistemas que sí la poseen. El segundo antecedente es el sistema SIAT, de administración tributaria, generado a través de un contrato de la municipalidad de Rosario con la empresa Tecso¹. Este sistema fue desarrollado inicialmente en los términos habituales de contratos de desarrollo, incorporándose posteriormente las adendas necesarias para garantizar su liberación bajo licencia GPL. El proceso de liberación se encuentra en desarrollo a la fecha. Respecto de este último caso, la diferencia principal estriba en que el proyecto GPLib tiene en su horizonte de trabajo el fomento para la creación y crecimiento de una comunidad de práctica alrededor del sistema. Esta comunidad será la base para la sustentabilidad a largo plazo de las innovaciones que deban incorporarse una vez concluido el contrato con la empresa proveedora.
- Está concebido como un espacio de colaboración, complementario al sistema en sí. Con esto se busca integrar a todos los actores que puedan aportar en la definición y construcción de la solución más adecuada. En ese sentido, se han seguido dos estrategias de integración en función del contexto:
 1. Interinstitucional. El proyecto no se inició con el llamado a licitación para su desarrollo. El hito fundacional consistió en la firma de un acuerdo marco entre la Facultad de Filosofía y Letras de la Universidad de Buenos Aires (FFyL-UBA) y la Facultad Latinoamericana de Ciencias Sociales (FLACSO). A partir del mismo, las sucesivas reuniones ayudaron a definir

1 http://www.rosario.gov.ar/sitio/gobierno/gestion/ciudad_digital7.jsp

desde el pliego licitatorio hasta el nombre del proyecto, contando con los valiosos aportes de los y las especialistas de ambas instituciones tanto en el campo de la bibliotecología como en informática. Además se buscó y obtuvo el involucramiento de las más altas autoridades de ambas instituciones.

2. Intrainstitucional. Todos los actores, desde los niveles de decisión pasando por la gestión intermedia y el personal operativo, fueron involucrados en el proceso de desarrollo. Para esto se habilitó un espacio en el gestor de proyectos de la FfyL (<http://proyectos.filo.uba.ar>) donde todos los usuarios, sin distinción de nivel de responsabilidad, recibieron una contraseña de acceso. En total, unas 30 personas obtuvieron el acceso a este espacio y es allí donde la ejecución de cada etapa es documentada. Para esto se convino que todas las solicitudes de información o definiciones que el proveedor necesitara fueron elevados en forma de “tickets”, espacios virtuales donde cualquier participante puede dejar su comentario, subir archivos que ejemplifiquen un proceso o simplemente dar su aprobación para la etapa en cuestión. Esto redujo el rol de “mediador” que la Dirección de Informática tiene en el proyecto, permitiendo no solamente un contacto más ágil con el proveedor sino un “alerta temprana” ante posibles funcionalidades que hubieran sido ignoradas en la etapa inicial de relevamiento.

- Es innovador desde el punto de vista de la tecnología de almacenamiento empleada, ya que recurre a CouchDB, una base de datos no relacional, orientada a documentos, cuya adopción crece geométricamente en los últimos años y es apoyada por Apache.
- Metodologías ágiles. Siguiendo la línea de trabajo de la Dirección de Informática de los últimos años, se ha optado por la utilización de metodología SCRUM modificada. Esto implica el trabajo en ciclos cortos, la liberación progresiva y temprana de los productos y la utilización de una herramienta que transparente todos estos procesos para todos los actores involucrados.
- Desde el punto de vista del lenguaje de programación empleado, python es el elegido por cada vez más instituciones públicas por su baja curva de aprendizaje, amplia disponibilidad de librerías de funciones y su entusiasta comunidad de desarrolladores que continuamente añaden nuevas funcionalidades y mejoras. Sobre este lenguaje se optó por utilizar el framework Django por su fuerte integración con CouchDB.
- Integra la mayor cantidad de soluciones posibles con el fin de disminuir los costos de desarrollo. En ese sentido, utiliza Greenstone como repositorio de documentos y herramienta de búsqueda full text y Moodle como interfaz para los usuarios finales que consumen servicios de la Biblioteca

(préstamos, solicitudes de apoyo referencial y otros).

- Está preparado para su uso en ambientes institucionales complejos, para lo cual incorpora en forma nativa la autenticación a través del sistema de directorio LDAP. Esta previsión permite que pueda utilizarse en forma independiente de otros sistemas o que pueda integrarse al directorio centralizado institucional, lo que permite a su vez que el usuario final pueda utilizar la misma contraseña que usa para otros sistemas como el e-mail, proxy ó red interna.

Beneficiarios

Los beneficiarios directos del proyecto serán todas aquellas agencias del área de gestión de conocimiento que encuentren en GPLib una solución para la gestión operativa y la gestión del conocimiento. En una primera etapa, antes de la liberación al público en general, los beneficiados serán las dos instituciones participantes en su elaboración. A partir de su publicación podrá ser utilizado por universidades, bibliotecas populares, escuelas y hospitales entre otros. También podrá tener amplia utilización en áreas no directamente vinculadas con la gestión de conocimiento, sólo como ejemplo éstas podrían ser los archivos de dependencias públicas o los centros de documentación y catastro de municipios.

En todos los casos, los beneficiarios indirectos serán los usuarios finales de los recursos bibliográficos y objetos de conocimiento (alumnos y lectores) y el personal afectado a las tareas de gestión que verá facilitado su trabajo diario.

Viabilidad Técnica, Financiera y Política Organizacional

El proyecto ya está en marcha y el plazo de ejecución del mismo ha sido fijado de antemano en el pliego licitatorio. Esto implica que existe una obligación contractual con el proveedor para entregar sus productos en tiempo y forma. Este factor es importante para entender que el ritmo de producción de resultados es constante y no tiene demoras sustantivas, ya que es el proveedor el mayor interesado en llevar a término el proyecto en el menor tiempo posible. El proyecto comenzó oficialmente en agosto de 2010 y en la actualidad el grado de avance se estima en un 60%.

En cuanto al aspecto político, es importante destacar que FLACSO tiene una política explícita de promoción del software libre, lo cual ha derivado, por ejemplo, en el plan de implementación de terminales Linux en todos sus puestos de trabajo. Desde la FFyL-UBA, el apoyo del Decano al proyecto y el papel fundamental de la Subsecretaria de Bibliotecas tanto en el apoyo político como en las definiciones técnicas ha resultado en el avance constante sobre los objetivos específicos planteados en el proyecto.

Finalmente, el alto grado de involucramiento del personal de las bibliotecas (personal en su mayoría no docente) es una fuerte garantía de continuidad del proyecto que va de la mano de las condiciones de continuidad de dichos puestos. Puesto en otros términos, el personal de estas áreas goza de una estabilidad y continuidad que supera los tiempos de gestión de los cargos electivos de ambas instituciones, y si el proyecto resuelve sus actuales problemas será sostenido con entusiasmo como ha quedado demostrado en el trabajo y las reuniones periódicas ya realizadas.

Ambiente de Hardware y Software

Los requerimientos de hardware son directamente proporcionales al tamaño de la instalación que se realice del mismo. En el caso de la implementación en la FFyL-UBA, se ha previsto el uso de una máquina virtual para el componente del servidor web que alojará la aplicación y la base de datos y otra para el alojamiento del componente Greenstone del sistema.

No es necesario adquirir ningún software adicional para la puesta en marcha y funcionamiento productivo de GPLib.

Relevancia para el Interés Público

GPLib influye en forma directa en el derecho a la educación. Al facilitar el acceso tanto a materiales físicos como digitales, y teniendo en cuenta los avances en otros planes nacionales como Conectar Igualdad ó Argentina Conectada que buscarán habilitar el acceso universal a Internet, el sistema tiene el potencial de transformarse en la puerta de entrada a vastos repositorios de documentación de todo tipo y materiales educativos de todos los niveles.

Como es sabido, la proliferación de herramientas de búsqueda ha convertido a las consultas simples (en Google ó similares) en la primer opción para adquirir información en Internet. La aplicación que estamos desarrollando puede ser un medio que contrapesa esta tendencia a la simplificación, un forma de que cualquier institución educativa pueda definir su voz dentro de la sobreabundancia de información disponible en la Red. Esta voz no necesariamente será la de las grandes universidades. En la medida en que el software puede implementarse con baja inversión y esfuerzo, más y más instituciones locales podrán utilizarlo para llegar a su propia comunidad en forma directa.

En definitiva, para que el derecho a la información pueda ser ejercido plenamente, deben facilitarse los medios para que los ciudadanos puedan acceder sin carga burocrática a los datos que buscan.

Facilidad de Reproducción

Las condiciones legales con las que el software será publicado permitirán que la puesta en marcha de una instancia del mismo sea tan simple como descargar una copia del código fuente desde el sitio dispuesto a tal fin e instalarla en un servidor propio.

Además de esta facilidad de reproducción, los bajos requisitos de hardware garantizan que todo el sistema pueda evaluarse en cualquier PC e inclusive utilizarse en el mismo entorno para la mayor parte de los usos posibles.

Finalmente, la proyectada existencia de una comunidad de práctica alrededor de la herramienta ofrece un panorama similar al de otros proyectos libres como el propio Greenstone, Moodle ó casos más conocidos como Apache ó el sistema operativo Linux. En todos estos casos, es la propia comunidad la que se da sus herramientas de autocapacitación, soporte comunitario, construcción y ampliación de documentación en forma colectiva. En este caso sin embargo, todas estas necesidades no quedarán únicamente cubiertas por el tiempo donado por voluntarios, ya que los actuales miembros del consorcio (como mínimo) aportarán todo lo producido para uso interno a la comunidad abierta, partiendo del principio rector de que **toda producción generada con fondos del Estado debería volver a los ciudadanos en todas las formas posibles.**