

Gestión consorcial en la implementación de repositorios

Rodrigo Díaz Pérez, Guillermo Toro Araneda y Cristóbal Caris Peñailillo
Alerta al Conocimiento S.A.
{rdiaz,gtoro,ccaris}@alerta.cl

RESUMEN

El desarrollo consorcial de repositorios tiene una serie de ventajas y beneficios. El consorcio Alerta al Conocimiento ha desarrollado una metodología para implementar y poner en operación repositorios académicos e institucionales de las universidades socias, bajo la cual se garantiza mantener la identidad de cada institución compartiendo el *know-how* tecnológico y disminuyendo los costos de desarrollo. Esto sobre una base segura, como es una plataforma computacional virtualizada.

INTRODUCCION

El consorcio de universidades chilenas Alerta al Conocimiento fue creado en 1993 para satisfacer necesidades de información de instituciones, empresas y personas, que realizan actividades académicas, de investigación y desarrollo en el país.

Actualmente integran el consorcio la Pontificia Universidad Católica de Chile, la Universidad de Concepción y la Universidad Católica del Norte.

La colaboración sistémica, tanto en adquisición cooperativa como en el desarrollo de proyectos tecnológicos ha beneficiado a las universidades socias y a sus clientes, produciendo economías de escala y transferencia de conocimientos.

Entre los desarrollos significativos del consorcio Alerta se encuentra el servicio **Hemeroteca Virtual** (1994), orientado a compartir digitalmente a nivel nacional las colecciones de publicaciones periódicas impresas, la liberación de la plataforma web de este servicio (2000) y el servicio **Metabuscador** (2005), que integra las colecciones de recursos electrónicos de las universidades socias en un espacio virtual colaborativo, manteniendo la identidad institucional.

En el 2009, Alerta pone en marcha el servicio **Repositorio Institucional** para las universidades socias.

Desde el 2011 el consorcio Alerta dispone de una plataforma computacional virtualizada como base para todos estos servicios, que forman parte de la oferta actual de Alerta.

La plataforma virtualizada hace posible que cada universidad disponga de un repositorio para desarrollo y pruebas, además del repositorio en producción.

La implementación consorcial de repositorios tiene múltiples beneficios. Entre otros, permite:

- Compartir información.
- Transferir tecnología.
- Concentrar el desarrollo en un equipo profesional especializado.
- Centralizar la administración y soporte técnico.
- Replicar cadenas productivas.
- Desarrollar buenas prácticas.

Realizar además esta implementación sobre una misma plataforma tecnológica simplifica la administración, garantiza la compatibilidad entre los repositorios y permite crear un repositorio consorciado, para dar aun más visibilidad a la producción académica de cada institución.

En el modelo de Alerta, las instituciones participantes delegan los niveles operativo y de explotación en el Centro Alerta, manteniendo en todo momento el control sobre la gestión y los contenidos de sus respectivos repositorios.

Cabe señalar que los servicios que implementa el consorcio benefician a una comunidad de 63.000 usuarios, compuesta por los académicos y estudiantes de los planteles mencionados.

IMPLEMENTACION CONSORCIADA DE REPOSITARIOS

En sus 19 años de existencia el consorcio Alerta al Conocimiento ha acumulado una valiosa experiencia en metodologías de desarrollo tecnológico.

Estas metodologías se han aplicado en el tratamiento de las colecciones y en el desarrollo general de funcionalidades en los sistemas consorciados.

El desarrollo del servicio repositorio sigue un modelo similar al de otros servicios implementados anteriormente en forma colaborativa.

Para poner en marcha un repositorio el consorcio Alerta, a través de su *staff* de profesionales, actúa como socio tecnológico de las universidades. Estas, a su vez, asignan un grupo de profesionales, que son los mandantes y quienes en última instancia deben validar los contenidos y funcionalidades del repositorio. Por lo común se trata de profesionales bibliotecarios e informáticos asociados a los sistemas de bibliotecas de las instituciones. Los profesionales del Centro Alerta conforman un equipo multidisciplinario que dialoga con los profesionales de las universidades e interactúa con ellos para definir cadenas productivas y realizar la implementación en general.

El desarrollo comienza con una primera fase, de Implementación, en la cual la idea es generar ciclos de desarrollo virtuosos caracterizados por una gran interactividad entre los profesionales del Centro Alerta y los profesionales de las universidades.

La fase de Implementación finaliza con la producción de una primera versión estable del repositorio, que es un primer punto de control a partir del cual se realiza el desarrollo evolutivo que tiene lugar en la siguiente fase.

Para controlar el proyecto, Alerta mantiene un documento maestro, que es una carta de navegación para el equipo de desarrollo y el equipo de profesionales que actúa como contraparte. En este documento, que es dinámico, se detalla las características técnicas de la instalación, las actividades realizadas y las actividades pendientes.

La siguiente fase, de Operación, implica un desarrollo constante a partir de la versión resguardada. En esta fase convive en paralelo el desarrollo con una versión en operación, que se mantiene disponible para los usuarios finales.



Fig. 2: Proceso evolutivo

Cada universidad define las políticas, contenidos y servicios de su repositorio, como también el desarrollo estratégico del mismo. El Centro Alerta, por su parte, mantiene operativo el repositorio y se hace cargo de desarrollar los nuevos requerimientos planteados por las universidades.

Los requerimientos se implementan en una instancia de pruebas idéntica a la existente en producción. De este modo los “clientes” pueden validar si la solución implementada es transferible a producción o aún requiere modificaciones. Por lo general los requerimientos agregan funcionalidades al repositorio y, en algunos casos, implican definir una nueva cadena productiva, desde el punto de vista de la gestión.

Las universidades tienen autonomía plena para orientar el desarrollo de acuerdo a sus intereses y capacidades de ejecución, como también el control absoluto sobre la interfaz gráfica. De este modo se preserva la identidad institucional y se hace transparente la tecnología que sustenta el servicio.

El Centro Alerta administra la plataforma computacional y realiza cargas masivas y respaldos de información a nivel de explotación de los recursos de hardware y software. Se encarga además, como hemos dicho, del desarrollo operativo y del funcionamiento del servicio, propiamente tal.



Fig. 1: Fases de Implementación y Operación

A nivel de operación se efectúan procesos y utilizan herramientas de software propias del funcionamiento del repositorio. Actualmente los repositorios funcionan sobre DSpace y se utiliza el entorno de desarrollo Eclipse para implementar las nuevas funcionalidades requeridas por cada universidad.

En todos los desarrollos, el hecho de contar con un servidor de pruebas homólogo al servidor en producción y funcionando sobre la

plataforma virtualizada facilita la experimentación y la comunicación con las universidades.

Al trabajar en un entorno consorciado, las soluciones tecnológicas adoptadas benefician a los diferentes actores, ya que los desarrollos son replicables y cualquier implementación exitosa puede ser sugerida a las demás instituciones, produciéndose un efecto de nivelación hacia arriba (*catch up*).

La implementación de nuevas funcionalidades, originadas en los requerimientos de cada institución, da lugar a la mejora continua en todas las instancias, lo que es un aspecto clave en el modelo de desarrollo de Alerta.

El procedimiento estándar ante los nuevos requerimientos es el siguiente:

1. El Centro Alerta realiza una evaluación técnica de cada requerimiento.
2. A partir de esta evaluación, hace una propuesta formal de mejora, dentro de la estructura general del proyecto de desarrollo o plan maestro.
3. Cuando la universidad acepta la propuesta y según las prioridades establecidas, el Centro Alerta realiza una implementación de prueba.
4. La universidad valida el desarrollo y este se incorpora a la nueva versión en producción.
5. El Centro Alerta informa la disponibilidad de esta nueva funcionalidad a las otras universidades, para que decidan si es adecuado transferirla a sus propios repositorios.
6. Si otra universidad considera que implementar la funcionalidad en su repositorio es apropiado, simplemente se interviene la última versión resguardada en el ambiente de prueba, para que la universidad valide la nueva mejora y acepte fijarla en el repositorio en producción.

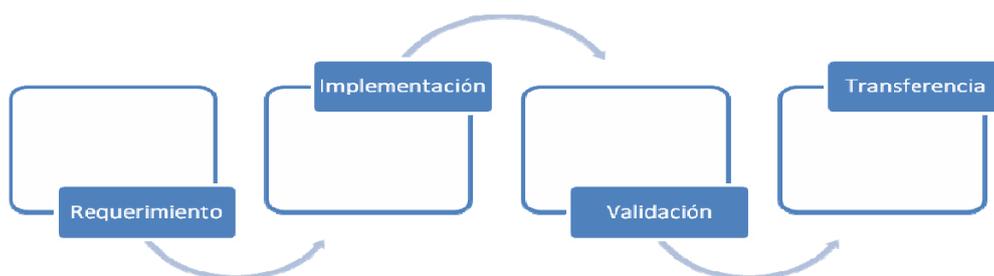


Fig. 3: Implementación de nuevos requerimientos

PLATAFORMA VIRTUALIZADA

Desde el año 2011 el Consorcio Alerta dispone de una plataforma computacional virtualizada, que está en la base del funcionamiento de todos los servicios.

Alerta optó por el uso de VMWare, considerando que es un sistema estable, compacto y con una gran base de usuarios a nivel mundial.

VMWare permite optimizar el uso de recursos computacionales y respaldar la imagen completa del servidor, lo que hace posible volver sobre *puntos de control* previos ante cualquier problema con los datos o los servicios.

Como ambiente de desarrollo, además, VMWare es un ambiente portable, lo que aumenta la seguridad en caso de dificultades mayores con una instalación. Encontrar proveedores que den servicio en esta plataforma es relativamente sencillo, por lo cual también se puede prevenir una catástrofe en el hospedaje actual manteniendo respaldos o transfiriéndolos a otro proveedor para activar el funcionamiento del servicio en cuestión de horas.

La plataforma virtualizada permite bajar los costos de hardware, ya que un servidor virtual ocupa solo una fracción de un servidor físico y la disminución de servidores físicos, además de ser un ahorro en sí misma permite disminuir los tiempos de administración y acortar los tiempos de implementación.

Por otra parte, en un servidor físico, los problemas técnicos a veces obligan a realizar una restauración desde cero, lo que claramente es una limitante para la experimentación.

En el caso del consorcio Alerta, disponer de esta plataforma ha permitido aplicar un método de desarrollo evolutivo con un ciclo en que se interviene la versión resguardada, se implementan y prueban los nuevos requerimientos y, cuando estos han sido validados, se fija una nueva versión que entra en producción con un nuevo nivel de madurez.

Este ciclo se puede desarrollar con efectividad y transparencia, dado que los repositorios virtualizados en el ambiente de desarrollo y en ambiente de producción son homólogos, por tanto los cambios validados se transfieren reemplazando la versión en producción con la nueva versión resguardada completa en cuestión de minutos.

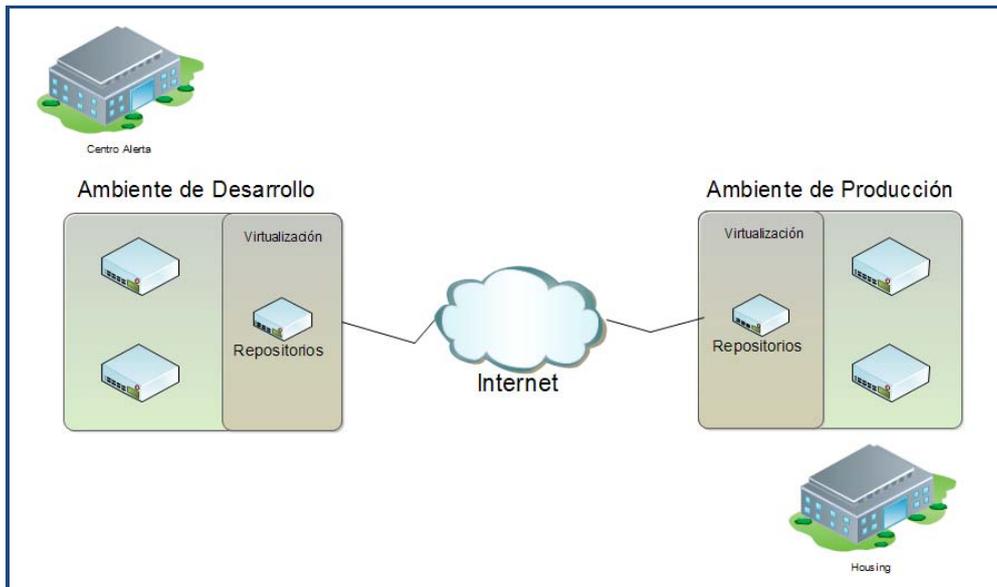


Fig. 4: Repositorios virtuales

Si la implementación llega a corromperse por cualquier motivo, se recurre a la versión resguardada original, para continuar probando y modificando el código y las parametrizaciones hasta lograr el nivel evolutivo esperado.

La virtualización ha permitido administrar en forma segura los procesos de instalación, configuración y modificación, a nivel de software, datos y recursos de información. La posibilidad de realizar "rollback" en forma integral hacia los puntos de control y mantener respaldos estructurados de los servicios con herramientas de amplio uso, como las provistas por el ambiente VMWare, permite dar garantías de alta disponibilidad, independencia del hardware y portabilidad de las aplicaciones, haciendo más seguro cualquier desarrollo y dando sustentabilidad a los servicios.

Producto de esto, Alerta ha mejorado considerablemente sus tiempos de respuesta, lo que se puede ejemplificar en el caso de la implementación tecnológica de un repositorio en la versión 1.8.2 de DSpace, realizada en un período de tiempo muy inferior al de las implementaciones anteriores.

IMPLEMENTACION DE DSPACE v. 1.8.2

La versión 1.8.2 de DSpace fue liberada el 24 de febrero de 2012 con el objetivo de reparar errores presentes en las versiones 1.8.0 y 1.8.1.

En el caso de Alerta, se optó por implementar la versión 1.8.2 debido

a las mejoras nativas de esta versión en cuanto al manejo de los diacríticos y a la fiabilidad en la indexación de los mismos.

La versión 1.6.2, utilizada como plataforma para los repositorios desarrollados por el consorcio Alerta requería parches para trabajar con los diacríticos. Los datos debían ser almacenados sin diacríticos y ser transformados al momento de la consulta.

Además de solucionar los problemas de tratamiento de los diacríticos y optimizar la indización, en la versión actual se cuenta con la inclusión nativa del tema Mirage, donado a DSpace por la empresa @MIRE. Este mejora la apariencia y percepción del usuario. Junto a esto, Mirage permite agregar filtros y sugerencias y facilita la personalización gráfica con CSS-3.

Otra mejora, que sin duda ayuda mucho a diseñar una mejor experiencia de usuario, es la incorporación de Discovery, como nuevo motor de búsqueda, paralelo a Lucene, a partir del cual se habilita una serie de filtros para consulta de información.

El hecho de que, adicionalmente, PostgreSQL ahora maneje la codificación de caracteres UTF-8, ha facilitado aun más el procesamiento de los diacríticos.

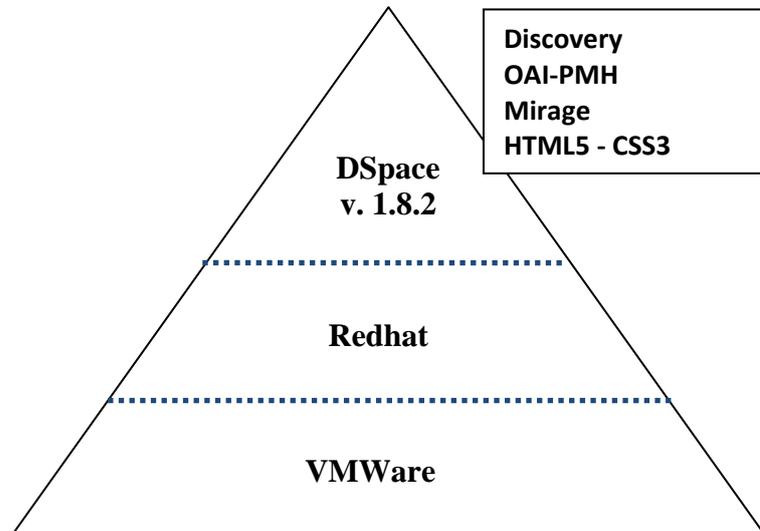


Fig. 5: Plataforma de software

La instalación actual de DSpace funciona sobre VMWare, RedHat y con las siguientes versiones de software utilitario:

- PostgreSQL versión 8.4
- Tomcat versión 6
- Ant versión 1.8
- Maven versión 2.2
- Java versión 6.32

Al nivel más bajo, de Explotación, el esquema de trabajo puede resumirse en la siguiente figura:

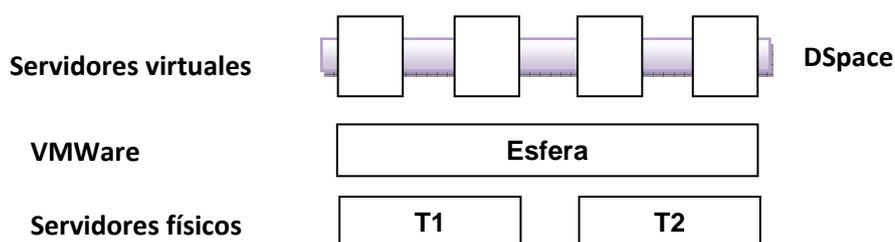


Fig. 6: Plataforma de explotación

CONTENIDO E IMPLEMENTACION DE LOS REPOSITORIOS

El servicio repositorio ofrecido por Alerta está en una etapa inicial en términos de contenido.

Actualmente están en producción el Repositorio académico de la Pontificia Universidad Católica de Chile (<http://repositorio.uc.cl>) y el Repositorio académico de la Universidad de Concepción (<http://udec.alerta.cl>), ambos sobre DSpace versión 1.6.2 y en proceso de migración o *upgrade* a la versión 1.8.2.

A la fecha, el repositorio de la Universidad Católica contiene 724 ítems y el de la Universidad de Concepción almacena 264 ítems, lo que representa un porcentaje mínimo en relación al potencial de crecimiento estimado para ambos repositorios. En cuanto a formatos almacenados, por el momento ambos solo contienen documentos pdf.

La Universidad Católica tiene colecciones institucionales compuestas por discursos de autoridades, publicaciones de académicos, publicaciones institucionales y tesis. Además tiene una colección patrimonial de libros digitalizados. La Universidad de Concepción solamente tiene tesis de postgrado y apuntes para uso docente.



Fig. 8: Repositorio UC en producción



Fig. 9: Repositorio UC en desarrollo

Las universidades, naturalmente, son las productoras de contenido y conforman colecciones que tienen un sentido local e institucional. Sin embargo, existe el interés de dar mayor y mejor visibilidad a algunos de

estos contenidos, por lo que el Consorcio Alerta dispondrá además de un repositorio institucional, denominado **Difunde**.

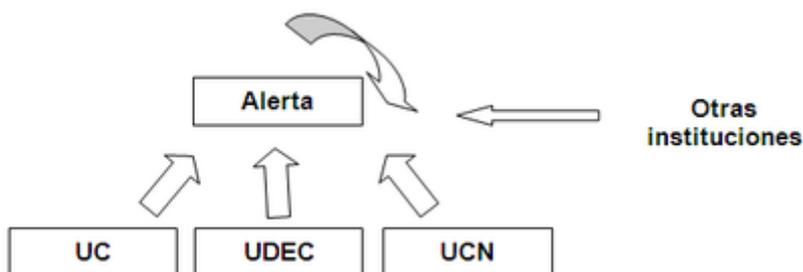


Fig. 7: Fuentes de datos/ítems para el repositorio consorcial Difunde

En este repositorio se cosechará datos e ítems provistos por las universidades que integran el consorcio, datos generados por la I+D de Alerta y datos recolectados eventualmente desde universidades e instituciones externas.

Para implementar los repositorios el Centro Alerta articula los requerimientos de las universidades e incorpora las ayudas de otros socios tecnológicos.

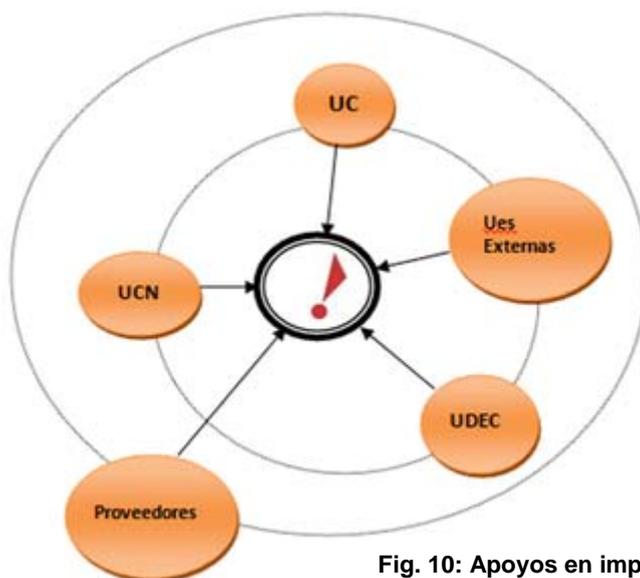


Fig. 10: Apoyos en implementación

En algunos casos los requerimientos presentan complejidades que deben ser resueltas con ayuda de instituciones externas o por miembros individuales de la comunidad de desarrollo de DSpace a nivel internacional. La Universidad Nacional de La Plata, en Argentina, ha sido

un colaborador permanente, con el cual el Consorcio pretende afianzar lazos de cooperación técnica a mediano plazo.

Dado que las universidades trabajan sobre la misma plataforma, muchos desarrollos pueden ser replicados, lo que trae beneficios institucionales mutuos, produciéndose sinergia y economías de escala, dado que se rentabiliza los desarrollos y produce transferencia de conocimiento.

Los proveedores de software básico (VMWare, DSpace, RedHat, etc.) y de aplicaciones (@MIRE y otros), también son una gran ayuda, sobre todo cuando han dedicado esfuerzos a documentar con claridad los procedimientos de instalación y parametrización.

El servicio repositorio se concibe en interoperabilidad con otros servicios del Consorcio Alerta, como el Metabuscaador implementado en cada universidad y con la Preservación Digital, que necesariamente será abordada en el futuro.

CONCLUSIONES

- El desarrollo consorcial de repositorios es una opción que presenta múltiples ventajas. Los participantes reciben beneficios cruzados y pueden mantener su identidad y desarrollo estratégico autónomo.
- La Innovación no reside en el *upgrade* de las tecnologías, sino en el desarrollo conceptual y en las metodologías de desarrollo de los proyectos.
- La virtualización de la plataforma base reporta múltiples ventajas. Entre otras, disminución de los costos de administración y hardware, sin embargo, lo fundamental es que se dispone de un ambiente de desarrollo seguro, que invita a la experimentación y el mejoramiento continuo.
- En Chile el desarrollo de repositorios institucionales y académicos aún es incipiente. La comunidad de desarrolladores es global, sin embargo tanto para las instituciones como para los profesionales sería beneficioso que hubiese más proyectos en el país. Un aspecto a desarrollar es la recolección automática de contenidos locales (*harvesting*).