

## **Propuesta de Base de Conocimientos para Proyectos Multidisciplinarios en el Área de Ingeniería Ambiental**

SALAS, Alfredo Cristóbal, GONZÁLEZ-ROCHA, Sergio Natán, LÓPEZ-HERNÁNDEZ, Ismael

A. Salas´, S. Gonzáles´, I. López´´

´Universidad Veracruzana, ´´Centro de Supercomputación de Barcelona.  
acristobal@uv.mx

Gonzáles - Hernández, María de los Ángeles, Domínguez - Basurto, Maribel, García-Durán, Atanasio. (eds.) *Educación Ambiental desde la Innovación, la Transdisciplinariedad e Interculturalidad*, Tópicos Selectos de Educación Ambiental-©ECORFAN-Veracruz, 2015.

## Resumen

Este artículo propone un sistema computacional basado en el diseño de un centro de datos para compartir información y documentos del área ambiental entre la comunidad universitaria. Este proyecto busca fomentar la participación de profesores y estudiantes en proyectos o trabajos de investigación multidisciplinarios. El sistema computacional incluye datos económicos, de salud, forestales, poblacionales, industriales y geográficos proporcionados por el Instituto Nacional de Estadística, Geografía e Informática del gobierno mexicano, datos atmosféricos y de contaminación ambiental del estado de Veracruz así como tesis, artículos científicos, trabajos de investigación de clase, etc. El sistema incluye distintas interfaces para acceso a la información: (a) servicios web para acceso a datos, (b) interfaz web para acceso georeferenciado de datos económicos, forestales, de salud, poblacionales, industriales (c) una interfaz semántica para inter-relacionar los conceptos con los documentos. Con este sistema se promueven los trabajos multidisciplinarios, la investigación se hace más robusta y completa a la vez que se tiene mayor material didáctico realista para analizar la situación ambiental actual de la región.

Base-conocimiento, proyecto, ambiental, Multidisciplinar, Colaboración

## Introducción

En los últimos años, se ha incrementado la publicación y el acceso a la producción científica-académica gracias a los avances en la comunicación digital y a la evolución de herramientas computacionales basadas en interfaces Web. Esta tendencia ha impulsado la creación de nuevas estrategias de publicación, modelos y herramientas para compartir información entre científicos y académicos. Ejemplos clásicos de esta idea son los sistemas IEEEExplore y ACM Digital Library los cuales son excelentes herramientas para compartir información relacionada con la academia en todo el mundo. Sin embargo, el acceso a información estratégica en estos sistemas no es de acceso libre sino que está restringido a suscriptores quienes pagan por dicho acceso. Esta política de acceso frena el desarrollo de la ciencia y la tecnología mundial sobre todo para individuos o instituciones de bajo presupuesto. Esta situación ha sido revisada en Benkler, Y. (2006), y en Björk, et. al. (2012), donde se analiza el impacto en las libertades de acceso a la información científica. Es por ello que este proyecto propone la creación de un repositorio de datos y documentos que sea abierto a la comunidad universitaria regional donde la información está inter-relacionada por varias disciplinas o áreas del conocimiento. Es parte de la estrategia que múltiples áreas académicas universitarias provean de información que posiblemente pueda relacionarse para generar nuevo conocimiento.

La idea de compartir la información no es nueva; en (MacGregor, 2014) se presenta el proyecto de conocimiento público el cual está activo desde 1998 donde se han construido herramientas como: Open Journal Systems, Open Conference Systems, Open Monograph Press con las cuales se pretende proveer de un camino alternativo para publicar material científico-académico y hacerlo disponible al público en general. Se reportan miles de profesores o investigadores quienes usan estas herramientas bajo la política de “acceso abierto”.

En Willinsky, (2005), Stallman, (2004) y en Paul, (2004), se presentan ideas sobre la evolución que podría tener el acceso abierto en la siguiente década en la cual ya estamos viviendo. En Ferreras-Fernández et. al, (2013), Karanastasis, (2014) y en Lindman, (2015), más de diez años después se siguen proponiendo plataformas ahora más sofisticadas para compartir información entre académicos. En este artículo se presenta una plataforma computacional con capacidad semántica y de red social y módulos que contemplan fuentes de datos de acceso abierto.

Incluso, en European Commission (2015), se emitió una convocatoria de apoyo financiero a la investigación de e-infraestructuras en donde se asegura que Europa requiere de políticas de acceso abierto basado en tecnología ya existente que permita el acceso confiable y permanente a registros científicos digitales. Se busca establecer un enlace entre la literatura y los datos para habilitar una evaluación transparente de la investigación y la reproducibilidad de los resultados. En esta convocatoria se presentaron las siguientes necesidades de innovación tecnológica: (1) Una e-infraestructura de datos dirigido por el servicio que responda a requerimientos específicos de investigadores y organizaciones para acceso abierto a depósitos de información científica. (2)

Desarrollo de conceptos probados y prototipos de nuevos servicios en soporte de ciencia abierta para asistir a profesores e investigadores en sus tareas diarias. (3) Soporte a interoperabilidad global de e-infraestructuras de acceso a datos abiertos y enlace de iniciativas similares alrededor del mundo para complementar el acceso físico a facilidades de investigación con acceso a datos para asegurar que Europa juega un papel de liderazgo in colaboraciones internacionales.

Por otro lado, la Fundación Nacional para la Ciencia o The National Science Foundation, (NSF, 2015) ha propuesto un plan para incrementar el acceso público a publicaciones científicas y datos científicos digitales resultantes de investigaciones patrocinadas por la fundación. El plan se titula “Today’s Data, Tomorrow’s Discoveries” y en él se plantea promover el compartir datos y resultados de investigaciones. La NSF requerirá que cada manuscrito: (1) sea depositado en repositorios de acceso público designado por la NSF. (2) Esté disponible para descarga, lectura y análisis libre de costo antes de los 12 meses desde su publicación. (3) Posea un mínimo de metadatos leíbles por una máquina, (4) sea administrado y preservado por largo tiempo. Esta política entrará en acción en todos los proyectos que sean patrocinados, sometidos o tengan fecha de vencimiento en Enero 2016. Toda esta información fortalece la idea de iniciar en la Universidad Veracruzana un repositorio de datos y documentos de acceso abierto para satisfacer la demanda de información requerida por estudiantes, profesores e investigadores universitarios. En la institución se hace evidente la falta de cultura de colaboración y compartición de datos entre estudiantes, maestros, investigadores universitarios. Incluso se fomenta el trabajo individual en la elaboración de tesis y trabajos de investigación por considerar que tiene mayor relevancia y se desestima el trabajo en equipo o incluso el trabajo multidisciplinario tampoco es promovido, y con esto se hace referencia a trabajos realizados entre profesores, investigadores o estudiantes de distintas disciplinas. Por ejemplo, un estudio de afectación a la salud de la población civil provocado por contaminantes ambientales emitidos por la industria petrolera de la región.

Es por ello que la idea de bases de datos abiertas al público debe ser promovida al interior de las instituciones de educación superior para acelerar la investigación científica y tecnológica en el país. En este artículo se propone una idea simple de realizar este proyecto que puede ser fácilmente replicable en todo el país. En la siguiente sección se mencionan detalles del diseño e implementación de este sistema computacional.

### **Diseño del sistema computacional**

El sistema computacional descrito en la anterior sección debe tener los siguientes requerimientos de software funcionales generales:

1. Administrar datos. El sistema debe poder almacenar, buscar y recuperar datos en distintos formatos como por ejemplo: datos climáticos y de contaminantes ambientales, enfermedades en la población civil geo-referenciadas, datos financieros, forestales, económicos, poblacionales, e industriales del estado de Veracruz.

2. Administrar archivos. El sistema debe poder almacenar, buscar y recuperar tesis, artículos científicos, trabajos de investigación de clase, etc.

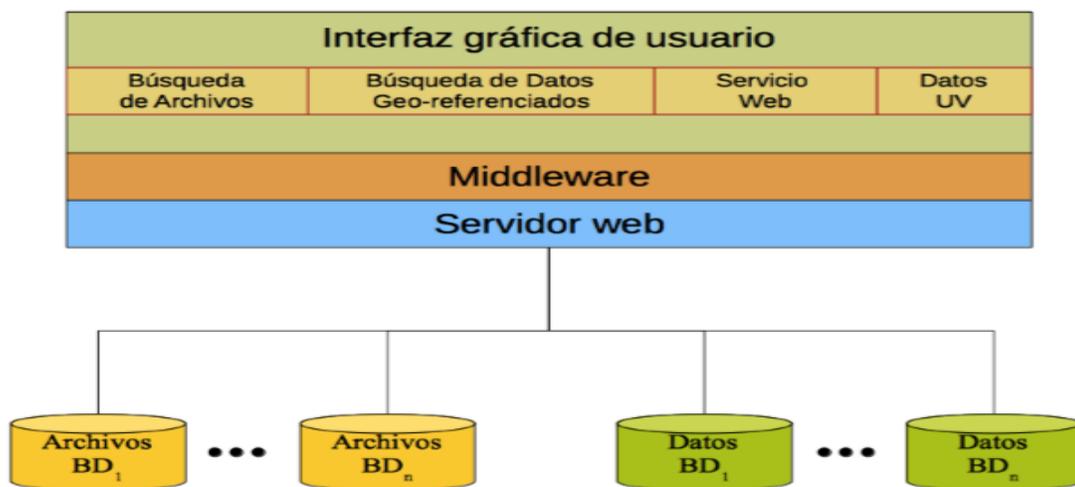
3. Acceso fácil e intuitivo a datos/archivos. Se prefiere el uso de interfaces web o de servicios web para acceso desde distintos tipos de plataformas como dispositivos móviles.

4. Almacenamiento escalable. Los repositorios de datos y documentos deben ser escalables, esto es, deben poder crecer en espacio de almacenamiento pero sin disminuir el rendimiento total del sistema computacional.

### Arquitectura del sistema

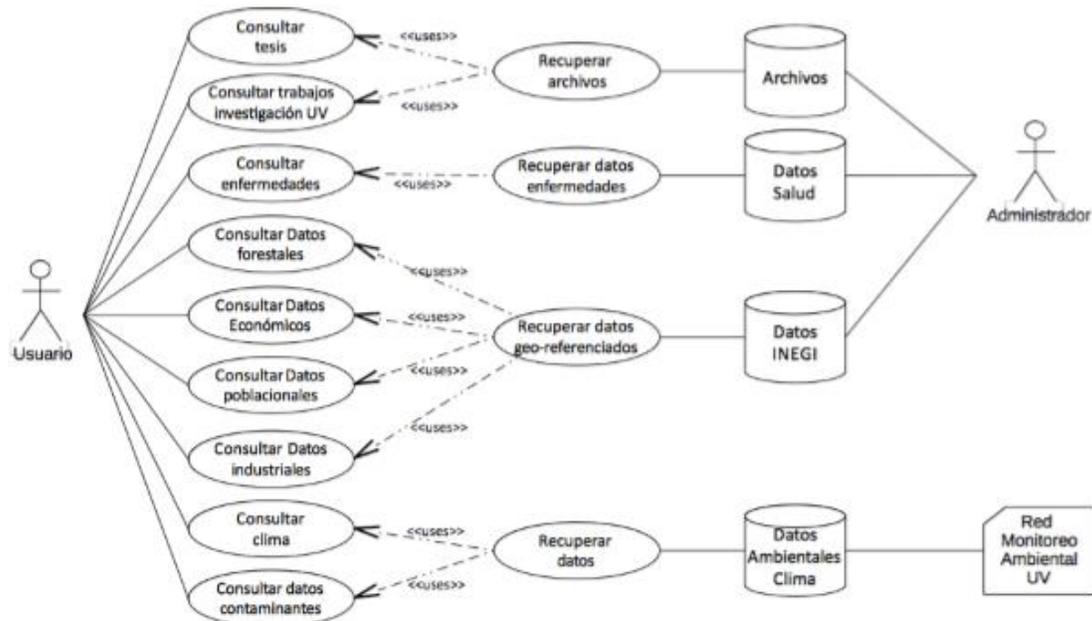
En esta sección se presenta la arquitectura del sistema computacional propuesto. En la Figura 1, se aprecia que el sistema contiene un almacenamiento incremental organizado en forma de bases de datos las cuales pueden estar geográficamente localizadas en lugares distintos. Cada base de datos contiene información referente a: datos climáticos y de contaminantes ambientales, enfermedades en la población civil geo-referenciadas, datos financieros, forestales, poblacionales, e industriales del estado de Veracruz.

**Figura 1** Arquitectura computacional del sistema que implementa la base de conocimientos



El sistema también considera un conjunto de bases de datos para almacenamiento de archivos organizados de manera semántica la cual es administrada por un motor de búsqueda de metadatos con enlaces a los archivos mediante una tabla de localización física de archivos. El sistema está organizado en tres capas de software: (1) Servidor Web. Donde se administran los servicios de transacciones bajo el protocolo de transferencia de texto (http). (2) Middleware. Donde se realizan las operaciones lógicas de direccionamiento y búsqueda de datos o archivos en las bases de datos. (3) Interfaz gráfica de usuario. En esta capa se tienen varias interfaces para búsqueda y despliegue de los datos y/o archivos. Entre las interfaces del sistema se tienen: búsqueda de archivos, búsqueda de datos geo-referenciados, servicios web para acceso a los datos a través de dispositivos móviles y datos estadísticos de la Universidad Veracruzana.

**Figura 2** Diagrama de diseño conceptual del sistema computacional que define la base de conocimiento UV



## Diseño conceptual

En la Figura 2 se muestra el diseño del sistema computacional en el cual se puede observar que el sistema tiene dos roles: (a) usuario final el cual tiene acceso a todo el sistema y (b) administrador el cual tiene bajo su responsabilidad resguardar y actualizar las bases de datos de salud, INEGI y los archivos generados por la misma institución. También, se tiene un tercer actor llamado “Red de Monitoreo Ambiental REMMA” el cual adquiere datos geoposicionados del clima y de contaminantes ambientales. Este actor es un sistema electrónico de telemetría que adquiere datos del estado de Veracruz de manera inteligente dependiendo de su geo-posición.

En la misma figura se puede notar que se tienen conceptualmente cuatro bases de datos: de archivos, de datos de salud, de datos del Instituto Nacional de Estadística, Geografía e Informática INEGI, Datos ambientales y clima. Estas bases de datos pueden ser una colección de bases de datos físicas dependiendo de la necesidad de almacenamiento y para conservar la escalabilidad (capacidad de crecer en almacenamiento sin alterar el rendimiento) del sistema.

## Implementación del sistema

En esta sección se presentan detalles de implementación del sistema computacional anteriormente presentado. El sistema computacional fue implementado en PHP con javascript y HTML versión 5. Las bases de datos están administradas por MySQL versión 5.6.2. La geoposición de los datos se hizo con la interfaz de programación aplicaciones de la empresa Google.

## Software utilizado

Joomla. Es un sistema de administración de contenido (CMS) el cual facilita la construcción de un sitio web junto con aplicaciones en línea. Un sistema de administración de contenido es un software que da seguimiento a cada pieza de contenido de un sitio web de la misma manera que una biblioteca da seguimiento a los libros que almacena. El contenido de este sitio web puede ser: texto simple, fotos, música, videos, documentos y cualquier otro archivo. Una de las principales ventajas de Joomla es que no se requiere de gran expertise para administrarlo dado que la mayor parte del trabajo lo hace el software. Este sistema de código abierto es flexible y fácil de usar además de extensible, esto es que puede agregarse funcionalidad mediante módulos que se instalan al sistema base. Este sistema está disponible libremente para todos por ser de código abierto.

OpenDocMan. es un sistema de administración de documentos (DMS) de código abierto, libre de uso y basado en Web el cual está escrito en PHP y cumple el ISO 17025 y el estándar para la administración de documentos. Este sistema tiene acceso basado en Web, control fino de acceso a los documentos e instalación automática de actualizaciones. OpenDocMan está desarrollado bajo la licencia GPL la cual permite modificaciones de manera libre.

OwnCloud. Este sistema permite el acceso universal a los archivos a través de una interfaz Web. Provee también de una plataforma para visualizar y sincronizar los contactos, calendarios y separadores de manera fácil a través de todos los dispositivos y habilita la edición básica directamente en la Web. La instalación requiere de requisitos mínimos del servidor y no necesita permisos especiales. El servidor OwnCloud es extensible vía un simple pero robusto Application Protocol Interface (API) para aplicaciones y plugins.

Open Journal System. Es un software de código abierto para la administración principalmente de revistas creado por el Public Knowledge Project, liberado bajo licencia pública general o GNU. Este sistema está diseñado para facilitar la apertura de publicaciones de acceso libre principalmente publicaciones con revisión entre pares. El sistema provee de la infraestructura de software que se requiere para la publicación en línea de artículos de revista sino que también aporta el flujo editorial completo incluyendo el envío de artículos, rondas de revisión entre pares y finalmente indexación de los artículos.

## Hardware utilizado

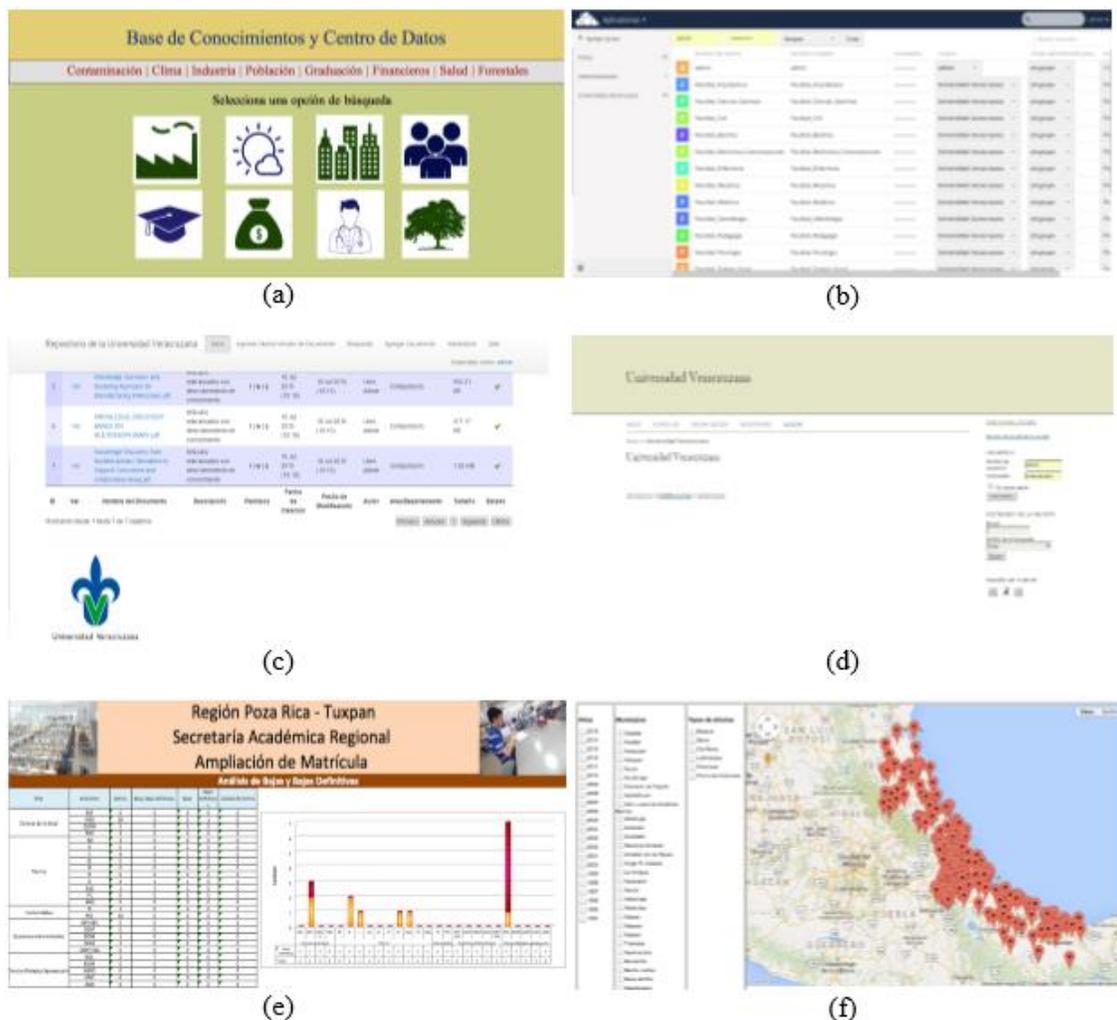
El sistema se ejecuta en una 8-core MACPRO 2.4Ghz, 12MB cache L3, 8GB SDRAM ECC, DDR2 1024Mhz, GDDR5 1Gb ATI Radeon HD 5870 corriendo sobre el sistema operativo OSX server 10.7.

## Resultados

En la Figura 3 se presentan capturas de pantalla de algunos de los módulos más importantes de la base de conocimiento. En la Figura 3a se muestra la portada del sistema donde se aprecian los módulos de monitoreo de contaminantes ambientales, monitoreo del clima, consulta de datos estadísticos de la industria, población, finanzas, y forestales provistos por el Instituto Nacional de Estadística, Geografía e Informática (INEGI) además de datos de salud y acceso a documentos académicos, tesis de grado y trabajos de investigación que genera la universidad. En la Figura 3b se muestra el módulo de administración de documentos OwnCloud donde se pueden compartir documentos entre los usuarios del sistema. Como se muestra en la figura es posible tener una interfaz basada en Web de común uso parecida a los administradores de documentos en los sistemas operativos.

Esta similitud a los sistemas de administración de archivos facilita la usabilidad del sistema entre los usuarios finales. En la Figura 3c se puede observar una captura de pantalla del módulo OpenDocMan el cual administra los documentos que se requiere de compartir entre los estudiantes, profesores e investigadores. En la Figura 3d se muestra una captura de pantalla del sistema de administración de revista en línea Open Journal System con la cual se pueden hacer publicaciones de artículos de investigación locales o regionales. En la Figura 3e se puede observar una captura de pantalla del módulo de análisis estadístico de la universidad con la idea de generar conocimiento de la misma institución. En este módulo se analiza el promedio y estado administrativo de estudiantes de una determinada matrícula. En la Figura 3f se observa una captura de pantalla del módulo de consulta de datos geoposicionados, en este caso los datos georeferenciados son datos forestales, de población, datos financieros y datos de salud cuyos datos son obtenidos del INEGI a través de su portal (<http://www.inegi.org.mx/>). En este portal se obtuvieron los datos estadísticos del estado de Veracruz.

**Figura 3**



*Figura 3.* Capturas de pantalla de la base de conocimientos de la Universidad Veracruzana. (a) Pantalla de inicio del sistema. (b) módulo OwnCloud, (c) módulo OpenDocMan, (d) módulo de Sistema de Revistas abiertas, (e) módulo de análisis estadístico universitarios, (f) módulo de datos georeferenciados.

Los datos de los módulos de consulta de datos climáticos, de contaminantes ambientales son obtenidos de la red de monitoreo ambiental de la Universidad Veracruzana. Este sistema obtiene los datos a través de micro-estaciones localizadas en las instalaciones educativas que transmiten en tiempo real a través del Internet a la base de datos central con resolución de toma de señales variable dependiendo del nivel de alertamiento de contaminantes o de islas de calor en la región.

Con los módulos instalados en la base de conocimiento podemos asegurar que se cumplen los requerimientos funcionales iniciales del proyecto puesto que se administran datos y archivos de manera integral además que se provee de interfaces gráficas principalmente basadas en Web donde los estudiantes, profesores o investigadores pueden utilizar sin necesidad de conocer aspectos técnicos de cada uno del software utilizado. Además, el uso de herramientas de código abierto permite el crecimiento escalar del sistema. Esto es, si la cantidad de datos o de documentos crece por encima de la capacidad inicial del sistema, es posible añadir módulo para administrar bases de datos remotas o cluster de bases de datos, esto es, repositorios remotos. Esto permitiría el crecimiento del sistema conforme se vayan incluyendo más datos con el transcurso del tiempo.

## Conclusiones

En este artículo se presenta un sistema computacional que integra múltiples módulos para consulta de datos y de archivos. El sistema en su conjunto también permite compartir archivos entre los usuarios de manera similar a un administrador de archivos. El sistema presentado cuenta con las siguientes características:

(1) Consultar datos geo-referenciados. El sistema permite la consulta de datos de: clima, contaminantes ambientales, salud, financieros, forestales, poblacionales, e industriales. Esto significa que el usuario final puede consultar datos de los temas anteriores y los resultados están organizados por municipio, por área o por estado. Esta característica permite tomar consciencia de la necesidad de equilibrio necesario para el desarrollo comunitario sustentable. Este indicador de sustentabilidad permitirá tomar mejores decisiones sobre crecimiento económico con menor impacto negativo en el medio ambiente.

(2) Compartir conocimiento. Uno de los problemas más recurrentes en las instituciones de educación superior en México es la falta de cultura para compartir el conocimiento logrado. Sin embargo, esta actitud no es por mala voluntad sino por falta de mecanismos para compartir, es por ello que este sistema permite que este objetivo de socializar el conocimiento sea alcanzable. De esta manera, si un estudiante quiere hacer una investigación de enfermedades presentes en un determinado municipio podría no solo concentrarse en las enfermedades sino que tendría a su disposición otros documentos o investigaciones realizadas en ese mismo municipio lo que permitirá robustecer la investigación inicial.

(3) Publicación arbitrada e indizada. Con el uso de herramientas para la publicación de revistas electrónicas es posible que de manera controlada se puedan publicar investigaciones realizadas por estudiantes, profesores o investigadores con la idea de que este conocimiento sea compartido entre la comunidad universitaria. Por ejemplo, si en algún municipio se hace un estudio de contaminantes en los mantos acuíferos esta investigación puede ser publicada al interior de la institución para que esta información pueda ser utilizada por otros universitarios para generar nuevas investigaciones como el impacto en salud de la población de estos contaminantes o el impacto en la agricultura de la región. La publicación de estas revistas internas puede aportar de manera significativa al conocimiento global multidisciplinario. El sistema presentado anteriormente hace uso de herramientas de código abierto que están disponibles sin costo.

El resto del sistema tiene una arquitectura de sistema de fácil re- implementación por lo que este artículo propone que esta investigación sea replicada en otras instituciones para analizar la evolución de la filosofía de bases de datos abiertas o compartidas, así como de la filosofía de ciencia de acceso abierto.

## Referencias

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*, New Haven, USA: Yale University Press. Disponible en: <http://goo.gl/irCc5d>. Última visita: 18 de julio del 2015.

Björk, B.-C., & Paetau, P. (2012). Open Access to the scientific journal literature—status and challenges for the information systems community. *Bulletin of the American Society for Information Science and Technology*, 38(5), 39–44. doi:10.1002/bult.2012.1720380512.

European Commission. (2015). Oportunidades de financiamiento. e-Infrastructures. H2020- EINFRA-2014-1. Disponible en: <http://goo.gl/YPntr8> . Última visita 18 de julio 2015.

Ferreras-Fernández, José A. Merlo-Vega, J-A., García-Peñalvo, F-J. (2013). Science 2.0 supported by open access repositories and open linked data. In proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality, Pages 331-332, ACM New York, NY, USA 2013. ISBN: 978-1-4503-2345-1 doi>10.1145/2536536.2536586.

Karanastasis, E., Andronikou, V., Chondrogiannis, E., Tsatsaronis, G., Eisinger, G., Petrova, A. (2014). The OpenScienceLink architecture for novel services exploiting open access data in the biomedical domain. In proceedings of the 18th Panhellenic Conference on Informatics. Pages 1- 6, ACM New York, NY, USA ©2014. ISBN: 978-1-4503-2897-5 doi>10.1145/2645791.2645811

Lindman, J., Kuk, G. (2015). From Open Access to Open Data Markets: Increasing the Subtractability of Open Data. In proceedings of the 2015 48th Hawaii International Conference on System Sciences. Pages 1306-1313, IEEE Computer Society Washington, DC, USA 2015. ISBN: 978-1-4799-7367-5 doi>10.1109/HICSS.2015.159

MacGregor, J., Stranack, K., Willinsky, J. (2014). *The Public Knowledge Project: Open Source Tools for Open Access to Scholarly Communication*. Stanford University, Stanford, USA. S. Bartling and S. Friesike (eds.), *Opening Science*, DOI: 10.1007/978-3-319-00026-8\_11, The Author(s) 2014

NSF. (2015). Public Access - To results on NSF-funded research. Disponible en: [https://www.nsf.gov/news/special\\_reports/public\\_access](https://www.nsf.gov/news/special_reports/public_access). Último acceso en 18 de julio del 2015.

Paul A. David, (2004). "Understanding the emergence of ‘open science’ institutions: Functionalist economics in historical context," *Industrial and Corporate Change*, volume 13, number 4, pp. 571–589. <http://dx.doi.org/10.1093/icc/dth023>

Richard Stallman, (2004). "Free software — free society! Interview with Richard Stallman (+ transcript)," *Indy Media UK*. Disponible en: <http://goo.gl/wATr39>, Último acceso en 18 Julio 2015.

Willinsky, J. (2005). The unacknowledged convergence of open source, Open Access, and Open Science. *First Monday*, 10(8). Available at: