

## Construcción de un Prototipo de Repositorio Temático en el campo de la Ingeniería de Software

María V. Doria<sup>1</sup>, Germán A. Montejano<sup>2,3</sup> & Claudia I. Inchaurredo<sup>1</sup>

<sup>1</sup>Dpto. Sistemas – Facultad de Tecnología y Cs. Aplicadas – Universidad Nacional de Catamarca

Maximio Victoria 55 - C.P: 4700 - San Fernando del Valle de Catamarca

TEL. 03833- 435112

vanesadoria@gmail.com, cinchaurredo@sistemas.frc.utn.edu.ar

<sup>2</sup>Facultad de Cs. Físico, Matemáticas y Naturales, Universidad Nacional de San Luis, Ejército de los Andes 950 – CP: 5700 – San Luis

<sup>3</sup>Facultad de Ingeniería, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa – La Pampa  
gmonte@unsl.edu.ar,

### Resumen

El presente trabajo tiene por objeto mostrar la línea de investigación que se llevará a cabo para la construcción de un prototipo de Repositorio Temático en el campo disciplinar de la Ingeniería de Software que pretende ser implementado en el Dpto. de Sistemas de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

El repositorio de información temático a construir debe permitir el acceso a producción científica de manera libre y abierta, para ello es importante identificar los pasos necesarios para su construcción.

Asimismo para el acceso a la información a través de repositorios de información se deben tener en cuenta las iniciativas que propone el Acceso Abierto (AA).

**Palabras Clave:** Acceso Abierto, repositorio temático.

### Contexto

El desarrollo y crecimiento de las estrategias que forman parte de las iniciativas de AA, representa un beneficio fundamental para la comunicación científica y académica mundial. Estas estrategias están referidas a los repositorios institucionales y temáticos, que almacenan la producción científica de instituciones científicas o académicas, constituyendo la vía más propicia para elevar la visibilidad y el impacto de la investigación. La literatura existente sobre desarrollo y gestión de repositorios temáticos es escasa comparada con la amplitud de información alcanzada por repositorios institucionales.

De acuerdo al contexto mencionado, con esta investigación se busca aportar información sobre los pasos a seguir para el desarrollo y gestión de un repositorio temático en el campo disciplinar de Ingeniería de Software.

La línea de investigación presentada en este trabajo se desarrolla en el marco de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis.

### Introducción

La comunicación científica es una de las instancias fundamentales de la actividad investigadora, siendo la publicación el mecanismo formal que regula el proceso por el cual los investigadores realizan sus contribuciones al acervo de conocimientos científicos. De este modo se genera la literatura científica que va configurando la dimensión bibliográfica de las disciplinas, entendidas como sistemas de producción y disseminación de conocimientos [1].

Un repositorio de información, institucional ó temático, representa una alternativa a la publicación y difusión de la literatura científica; si bien no es un canal formal de publicaciones sirve como estrategia para acceder a la información de manera libre y abierta. Libre se refiere a la consulta del documento completo en línea sin costo y abierta esta relacionada con algunos derechos de uso adicionales como el de modificar y distribuir la obra siempre y cuando se cite al o los autores.

Enfocándonos en el objeto de estudio de esta investigación, los repositorios temáticos, existe un consenso general respecto a que fueron los

primeros repositorios en aparecer, pueden describirse como proveedores de servicios que recolectan datos estructurados relacionados con un tema, permitiendo posteriormente al usuario buscar a través de muchas fuentes distribuidas en todo el mundo desde un punto de acceso único [2][3].

Surgieron en el ámbito de disciplinas académicas concretas, tales como, la física, ciencias de la información, ciencias cognitivas, salud, entre otras.

Las características principales de los repositorios temáticos, son las siguientes [4]:

- Permiten almacenar documentos pre-prints (antes de pasar cualquier proceso de arbitraje) o post-prints (documentos revisados por pares y aceptados, publicados o en proceso de publicación) y pueden ser artículos de revistas, comunicaciones en congresos, capítulos para libros o cualquier otra forma de comunicación científica.
- Los documentos disponibles en los repositorios de información se encuentran a texto completo y el acceso es gratuito, libre de cualquier tipo de restricciones.
- Los propios autores son los responsables de la introducción de los documentos en el repositorio, mediante el proceso de autoarchivo.

El funcionamiento de un repositorio temático consiste en que el autor se registre y de esta manera se le otorga un espacio desde el cual incorporar los documentos. En este espacio introducirá los metadatos, descripciones estructuradas y opcionales que están disponibles de forma pública para ayudar a localizar el recurso (documento), es la información no relevante para el usuario final pero de suma importancia para el sistema que administra los documentos. Los metadatos son requeridos a través de formularios, y finalmente depositará el recurso, que según el mecanismo de publicación podrá visualizarse públicamente.

Los metadatos solicitados para el recurso son los que pertenecen a la clase de metadatos descriptivos, en el caso de trabajos científicos son título, palabras clave, resumen, fecha, entre otros que permiten la identificación y su localización a través de la computadora. Estos datos son recolectados por diferentes

proveedores de datos (administradores de contenido) que soportan el protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). [5]

Las características de este protocolo son [6]:

- Simplicidad: los creadores buscaron una forma simple que se encuentre al alcance de cualquier usuario.
- Normalización: basados en estándares ampliamente utilizados como el protocolo http (Hiper Text Transport Protocol) para la transmisión de datos y órdenes y XML (Extended Markup Lenguaje) para la codificación de metadatos.
- Recolección: existe una entidad que pone a disposición de los interesados información bibliográfica sobre los documentos que almacena.

Por otro lado, los elementos del protocolo son los siguientes [12]:

- RECURSO: Es el documento o producción científica
- ITEM: Colección de propiedades sobre un RECURSO, es decir, metadatos que describen el recurso (título, autor, etc.)
- REGISTRO: Es un ITEM expresado en un formato específico formado por:
  - Encabezamiento: contiene un identificador único (identifica de forma inequívoca un artículo dentro de un repositorio), marca de fecha (fecha de creación, modificación o eliminación del registro con el fin de la recolección selectiva) cero o el conjunto de miembros de la partida con el propósito de la recolección selectiva, un opcional de estado de un atributo con un valor de borrado que indica el retiro de la disponibilidad del formato de metadatos especificado para el elemento.
  - Metadatos: se debe especificar le modelo de metadatos elegido
  - Acerca de: es una parte opcional en la que se pueden definir declaraciones de derechos y declaraciones de procedencias.

El protocolo OAI-PMH recomienda usar el modelo de metadatos Dublin Core sin calificar. [7].

Este modelo se compone de 15 elementos agrupados según:

- Contenido
  1. Title (título): cuando el recurso no se trata de un documento de texto el título no es obvio, como el caso de imágenes, es posible utilizar una frase descriptiva.
  2. Subject (tema): incluye tanto el resumen como las palabras clave del recurso.
  3. Description (descripción): debe dar cuenta del contenido del recurso, por ejemplo puede incluir la introducción, una tabla de contenido o un texto descriptivo del recurso.
  4. Source (fuente): es un elemento relacionado a una Referencia a un recurso de la cual el propio recurso se deriva.
  5. Language (lenguaje): especifica el lenguaje del recurso.
  6. Relation (relación): referida con la relación que puede existir con otros recursos.
  7. Coverage (cobertura): relacionada con incluir la ubicación geográfica y duración temporal del recurso.
- Propiedad Intelectual
  8. Creator (autor): referido a las personas que tienen la propiedad intelectual del recurso.
  9. Publisher (editor): es el encargado de la edición que puede ser una persona o empresa.
  10. Other Agent (otras colaboraciones): se refiere a aquellas personas que sin ser autor/es o editores mantienen una relación directa con la realización del recurso.
  11. Rights (derechos): elemento referido a la propiedad intelectual, derechos de autor y licencias.
- Instanciación (Representación)
  12. Date (fecha): fecha a partir de la cual el objeto está disponible, la forma recomendada se define en la norma ISO 8601: AAAA-MM-DD.
  13. Type (tipo de recurso): es un elemento que sirve para especificar el género del recurso y su propósito.
  14. Format (formato): referido al formato del recurso o dimensiones (tamaño y duración).

15. Identifier (identificador): es un elemento para identificar el recurso por medio de una cadena conforme a un sistema de identificación formal, como un ISBN.

Otro tema importante en el ámbito de los repositorios temáticos es el conocimiento de los aspectos legales referidos a la propiedad intelectual, los derechos de autor y las licencias.

Sobre propiedad intelectual se debe conocer lo establecido en Argentina por la ley de Propiedad Intelectual 11723 [8] y lo establecido por la Organización Mundial de la Propiedad Intelectual (OMPI) [9].

La OMPI también establece los aspectos legales para el derecho de autor y las licencias.

### Líneas de Investigación y desarrollo

Para llevar a cabo la investigación la misma se dividió en fases:

- Fase Exploratoria: se ha analizado el objeto de investigación a partir de la información recopilada que procede no sólo de documentos tradicionales como artículos, tesis y libros sino también, por diversos recursos en línea como blogs, wikis, sitios Web de organizaciones e iniciativas de AA, entre otros. También han sido de especial utilidad otros tipos de documentos como informes o estadísticas nacionales e internacionales, que permitieron profundizar en el estado del arte.

El resultado obtenido es que a pesar que los repositorios temáticos son los primeros en aparecer, la literatura que discute el tema es limitada y se encuentra enfocada más hacia el campo de la Biblioteconomía y la Ciencia de la Información que a la construcción de repositorios temáticos.

- Fase Descriptiva: se han estudiado y definido las directrices de creación de repositorios de información.
- Fase Proyectiva: se ha comenzado el abordaje de la propuesta dentro de lo que se conoce como Investigación Aplicada ya que se propone transformar el conocimiento puro adquirido, en conocimiento utilizable, para ello se construirá un prototipo de repositorio

temático en el campo disciplinar de la Ingeniería de Software.

La construcción del prototipo del repositorio temático en IS se hará conforme a los siguientes pasos:

- Paso 1: Análisis de plataformas de software para la creación de repositorios de información. Se ha comenzado a trabajar en este paso.
- Paso 2: Selección de la plataforma de software adecuada para el repositorio temático en IS.
- Paso 3: Diseño del repositorio temático.
- Paso 4: Prueba del prototipo.
- Paso 5: Presentación del prototipo desarrollado a expertos y potenciales usuarios para conocer su opinión y reconocer las ventajas y desventajas del prototipo creado.

### Formación de Recursos Humanos.

Los estudios realizados en esta línea de investigación pertenecen al trabajo de Tesis de Maestría “*Modelo de un Repositorio en Ingeniería de Software para la Facultad de Tecnología y Cs. Aplicadas de la UNCa*”.

El equipo de trabajo está formado por:

Tesista: Lic. Maria Vanesa Doria.

Director: Mg. Lic. Germán Antonio Montejano.

Co – Director: Mg. Ing. Claudia Inés Inchaurredo

Con este trabajo de tesis se prevé continuar con la formación y fortalecimiento en investigación de la tesista a través de la guía del director y codirector.

Como así también poder interactuar con otros investigadores del país que se encuentran en la misma línea de investigación.

### Referencias.

[1] S. Miguel; El acceso abierto a las publicaciones de la comunidad científica argentina en el campo de la medicina. *Renata*, Vol. 1 N° 2, ISSN 2027-7415, pp 50-65, (2011)

[2] Berlin Declaration on Open Access to

Knowledge in the Sciences and Humanities. (2003) [online] <<http://oa.mpg.de/openaccessberlin/berlindeclaration.html>> [Consulta: 18 de marzo de 2011].

[3] Declaración de Salvador sobre acceso abierto: la perspectiva del mundo en desarrollo. [online] <[http://www.bvs.org.ar/pdf/Salvador-Acceso Abierto.pdf](http://www.bvs.org.ar/pdf/Salvador-Acceso%20Abierto.pdf)> [Consulta: 1 de Noviembre de 2011].

[4] Vásquez, G.; “Repositorios Digitales de Documentos”, Primer Curso sobre Escritura Científica y Repositorios de Acceso Abierto, 6 y 7 de diciembre de 2007

[5] Doria, M., acceso Abierto y Repositorios Temáticos, Editorial Académica Española, pp.9 ISBN 978 – 3-659 -06908 – 6, (2013)

[6] Y. Sarduy Domínguez, P. Urra Gonzalez, Herramientas para la creación de colecciones digitales, *Revista Acimed*, Vol. 14 N° 5, (2006) [online] Disponible: <[http://bvs.sld.cu/revistas/aci/vol14\\_5\\_06/aci19506.htm](http://bvs.sld.cu/revistas/aci/vol14_5_06/aci19506.htm)> [Consulta: 1 de Noviembre de 2011]

[7] Dublin Core [online] Disponible en: <http://dublincore.org/> [Consulta: 23 de Marzo de 2012]

[8] Ley de Propiedad Intelectual, [online] Disponible: <http://www2.mincyt.gov.ar/11723.htm>, [Consulta: 1 de Noviembre de 2011]

[9] Organización Mundial de la Propiedad Intelectual, ¿Qué es La propiedad intelectual?, OMPI N° 450(S). ISBN 92-805-1157-0