

Hacia el desarrollo de un repositorio institucional en UTN-FRSF: Aspectos técnicos a considerar

Fernanda Golobisky, Mariel Ale, Luciana Ballejos, Milagros Gutiérrez

CIDISI

Universidad Tecnológica Nacional – Facultad Regional Santa Fe

Lavaisse 610

{mfgolo, lballejos}@santafe-conicet.gov.ar, {male, mmgutier}@frsf.utn.edu.ar

Abstract

La creación de un repositorio institucional de acceso abierto para objetos digitales tiene como fin el difundir y discutir en la comunidad educativa de esta región los nuevos conceptos que se vienen gestando a fin de crear conciencia de la importancia de contar con una sociedad del conocimiento libre y plural, de fomentar el libre acceso, la reutilización, la visibilidad y la preservación de la propiedad intelectual universitaria en todas sus dimensiones. Uno de los problemas que se presentan en este tipo de sistemas es la interoperabilidad a nivel semántico, si bien el Sistema Nacional de Repositorios Digitales (SNRD) introduce lineamientos a cumplir en la implementación para poder ser cosechado, esto no es suficiente dado que los distintos repositorios pueden utilizar distintos estándares de metadatos. Este trabajo propone un marco conceptual para la implementación de repositorios institucionales en el contexto de universidades públicas, poniendo énfasis en la interoperabilidad semántica de la información contenida en los mismos. Se presenta desde un punto de vista tecnológico la implementación de un prototipo en la UTN – FRSF.

1. Introducción

El acceso a la información científica se ha convertido en un reto importante en el mundo actual debido a los altos costos que ello implica. Desde el Ministerio de Ciencia y Tecnología de la Nación se promueve la creación y uso de los repositorios institucionales en Universidades y centros de investigación del país.

La creación de un repositorio institucional de acceso abierto para objetos digitales educativos (ODE)[1] en la Universidad Tecnológica Nacional Facultad Regional Santa Fe (UTN – FRSF) tiene como fin el difundir y discutir en la comunidad educativa de esta región los nuevos conceptos que se vienen gestando a fin de crear conciencia de la importancia de contar con una sociedad

del conocimiento libre y plural, de fomentar el libre acceso, la reutilización, la visibilidad y la preservación de la propiedad intelectual universitaria en todas sus dimensiones.

La información juega un rol importante en esta era digital que estamos viviendo. Las universidades y centros de investigación generan mucha información que debe ser organizada y almacenada para su fácil y rápido acceso. Bajo este contexto aparecen nuevos conceptos relacionados: repositorios, metadatos y estándares.

Un repositorio es un sistema de información que organiza, almacena y brinda el acceso libre a la producción de la información [2]. Un repositorio en el contexto universitario contendrá objetos digitales educativos (ODE), siendo éstos cualquier recurso que pueda ser utilizado para la educación. Existen diferentes tipos de ODE que pueden ser compartidos y una gran variedad de estándares propuestos para la descripción de los mismos. Esta diversidad afecta sensiblemente la interoperabilidad de la información y particularmente su búsqueda, recuperación y reuso. Entre los estándares a considerar se encuentran Dublin Core (DC) [3] que utiliza categorías para describir el contenido y que fue creado para describir recursos web en general (título, autor, palabras claves, idioma, etc.), y Learning Object Metadata (LOM) [4] que permite describir aspectos educativos asociados al ODE. (nivel educativo, complejidad, audiencia esperada, etc.). La mayoría de los repositorios según se analizó en [5] utilizan DC. También el Sistema Nacional de Repositorios Digitales (SNRD) [6] en su recomendación, se basa en este estándar. Sin embargo, el mismo deja de lado aspectos pedagógicos que es importante considerar en la descripción de un ODE sobre todo si se pretende que el mismo pueda ser reutilizado, y que es fundamental considerar en ambientes universitarios. Son cada vez más las instituciones que se dan cuenta de esta problemática e intentan solucionarla agregando nuevos metadatos en la descripción de sus ODE sin perder por ello la interoperabilidad. A su vez, si bien la incorporación de

nuevos metadatos trae ventajas en la búsqueda y recuperación, trae aparejado la mayor carga de trabajo que esto significa para quien deposita material en el repositorio institucional (RI).

Este trabajo se enfoca en el aspecto tecnológico de la implementación de un repositorio, brindando una solución que facilite la descripción adecuada de los ODE, proponga mecanismos para facilitar el llenado de los metadatos y contribuya a la interoperabilidad semántica y sintáctica con otros repositorios.

Este trabajo se organiza como sigue. En la sección dos se presenta el marco conceptual sobre el cual se desarrolla el artículo. En la sección tres se presentan los avances y propuestas sobre distintos aspectos en la implementación del repositorio. En la sección cuatro, se muestran los resultados parciales obtenidos. Finalmente en la sección cinco se concluye el trabajo.

2. Marco conceptual de un repositorio institucional

Para la implementación de un repositorio institucional en una universidad pública, se identificaron distintos aspectos a considerar como se muestra en la Figura 1.

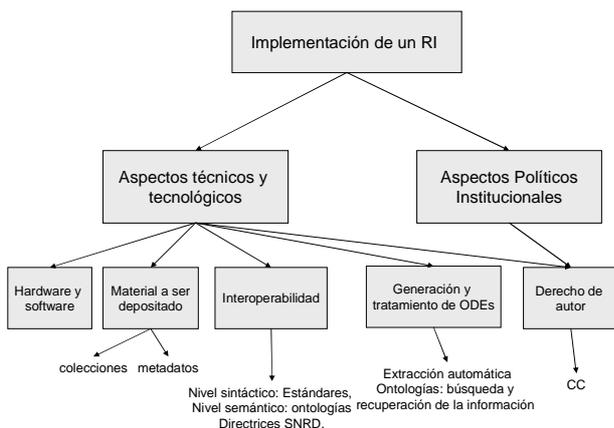


Figura 1. Aspectos en la implementación de un RI.

Este trabajo se concentra en los aspectos tecnológicos relacionados a la implementación. Como el gráfico muestra, hay cinco cuestiones a tener en cuenta desde el punto de vista tecnológico. (i) El hardware y el software que se van a utilizar y que son necesarios para cumplir con niveles de calidad del servicio que se quiere brindar. En [7] se presenta el análisis de las herramientas utilizadas y se desarrolla la implementación del prototipo identificando debilidades y fortalezas como también los trabajos a futuro. (ii) Análisis del material a ser depositado en el repositorio, donde es necesario determinar cuales son los tipos de ODE que el RI va a contener y en base a esta clasificación poder determinar

los metadatos que mejor describan a los objetos de estas colecciones. Para ello es necesario también hacer un análisis de los metadatos que pueden ser utilizados dependiendo no sólo de los estándares existentes sino también de las características propias de la plataforma de desarrollo seleccionada para implementar el RI. (iii) La interoperabilidad es la característica más importante sobre la que descansa el principal objetivo del AA: permitir el libre acceso a las publicaciones científicas en forma gratuita. El valor de estos repositorios se sustenta en la posibilidad de interconexión entre ellos, de manera de poder crear redes de repositorios y de esa forma aumentar la visibilidad, la reutilización de los recursos que en ellos se encuentran no solo por humanos sino también por máquinas [8]. Se consideran dos niveles, el sintáctico y el semántico. Varias iniciativas comienzan a surgir en el contexto de interoperabilidad, entre ellas es posible mencionar la iniciativa de archivos abiertos (OAI) que propone un protocolo para el intercambio de archivos, SWORD que propone un protocolo para estandarizar el depósito de material en los RI. Estas propuestas contribuyen a la interoperabilidad sintáctica permitiendo que los sistemas puedan compartir datos en el mismo formato pero no garantizan una correcta interpretación a nivel del significado de lo que se está intercambiando. (iv) La generación y tratamiento de un ODE involucra cuestiones relacionadas a la carga o depósito del material en el repositorio. También es necesario tener en consideración la carga de los metadatos. Una buena solución para facilitar la carga de los mismos es un extractor automático. La plataforma de desarrollo utilizada para la implementación del RI, especifica por defecto un conjunto de pasos a seguir, sin embargo este puede ser modificado si fuera necesario. Para la recuperación de un ODE entran en juego los metadatos definidos, el protocolo de comunicación utilizado para la interconexión con otros repositorios y también cuestiones relacionadas a cómo implementar algoritmos de búsquedas eficientes y eficaces. Finalmente el último punto a considerar desde la perspectiva de los aspectos tecnológicos, es lo referente a derecho de autor. Es importante definir en este tipo de RI que derechos se ceden y cómo se lo realiza. Ceder un derecho no significa renunciar al mismo, el acceso abierto no es equivalente a no reconocer autoría. Si se piensa desde el punto de vista tecnológico, esto simplemente implica la asociación de un archivo conteniendo los derechos a un ODE. Pero es importante identificar cuestiones políticas de cómo se respetan esos derechos.

3. Avances y propuestas para la implementación del RI – FRSF

Para el desarrollo del prototipo se trabajó en un equipo interdisciplinario atacando los distintos aspectos que fueron identificados en la sección anterior. Se mostrará a continuación la integración de los avances y las propuestas que se fueron desarrollando en los distintos grupos de trabajo. Para la implementación del repositorio institucional de acceso abierto para ODE en la UTN-FRSF se utiliza uno de los repositorios de código abierto más importante y usado en el ámbito académico, DSpace¹.

3.1. Especificaciones técnicas y características más importantes de DSpace

DSpace es una plataforma de software de código abierto que permite a las organizaciones:

- Capturar y describir material digital usando un módulo de presentación del flujo de trabajo.
- Distribuir los recursos digitales de una organización sobre la web a través de un sistema de búsqueda y recuperación.
- Almacenar y preservar contenidos digitales por largos períodos de tiempo.

La información que almacena DSpace está estructurada en 6 componentes básicos (ver Figura 2): 1) *Community*: este componente representa las comunidades y subcomunidades que pueden ser definidas en un RI para identificar conjunto de colecciones; 2) *Collection*: representan las colecciones, que como puede observarse en el gráfico, están conformadas por ítems que están relacionados entre sí ya sea por tener un tratamiento o flujo de trabajo en común; 3) *Ítem*: contiene los metadatos asociados a un ODE en particular, un ítem tiene asociado un *bundle*, que representa uno o mas archivos físicos que conforman el ODE; 4) *Bundle*: es un paquete de archivos, pueden ser archivos originales, de licencia, o generados a partir de los originales, entre otros; 5) *Bitstream*: es un archivo; 6) *Bitstream Format*: identifica el formato de un archivo, así un archivo tiene asociado un formato único.

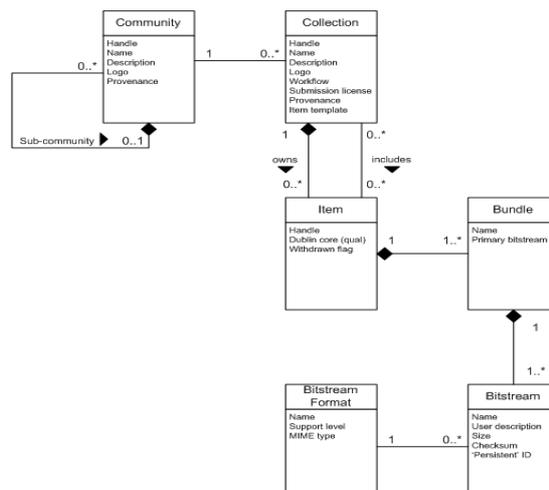


Figura 2: Modelo de datos de DSpace

Con este modelo la estructura de la organización que usa DSpace se ve reflejada en la manera en la que se organiza la información dentro de la plataforma. Cada sitio que utilice DSpace está dividido en comunidades, las cuales pueden ser divididas a su vez en subcomunidades que reflejan la estructura de la facultad, el departamento o centro de investigación de una universidad. Las comunidades contienen colecciones, que son agrupaciones de contenidos relacionados. Estas colecciones pueden aparecer en más de una comunidad. Cada colección está compuesta por ítems, los cuales son los elementos básicos del archivo. Cada ítem es propiedad de una colección. Cada ítem puede aparecer también en otras colecciones, aunque solo puede pertenecer a una colección. Cada ítem tiene un registro de metadatos Dublin Core. Los ítems pueden estar subdivididos en paquetes de bitstreams (archivos), es decir, los bitstreams que de alguna manera están relacionados, como por ejemplo los archivos HTML y las imágenes que componen un documento HTML, están organizados en paquetes. Cada bitstream está asociado con un formato de bitstream, correspondiente a cada uno de los formatos de los archivos que los usuarios envían. Cada formato de bitstream tiene, además, un nivel de compatibilidad (*soportado, conocido, no soportado*), indicando si la organización será capaz de preservar el contenido en ese formato en el futuro, implementado por el atributo *support level* de la clase *Bitstream Format*.

El modelo de datos de Dspace está soportado por el motor de base de datos PostgreSQL. Se trata de un motor relacional, poderoso y robusto, que permite almacenar toda la información (organización del contenido, metadatos, información del usuario, etc.) y mantiene índices que posibilita a los usuarios realizar las búsquedas normales de SQL.

¹ www.dspace.org

Los contenidos de los ítems se almacenan en directorios de un sistema de ficheros denominado *assetstore*. La referencia de un ítem a sus ficheros se encuentra en el campo *Internal_id* de la tabla *bitstream*, el cual contiene 38 dígitos. Este campo es usado para determinar la ubicación exacta de almacenamiento del archivo (Figura 3).

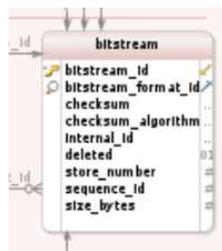


Figura 3: Estructura de la tabla bitstream

Así, por ejemplo, un bitstream con el identificador 11083282628192407436799614057093114020, indica que el correspondiente fichero (bitstream) se encuentra buscando los seis primeros dígitos del identificador, que indican en que subdirectororio de tercer nivel está el ítem (11 >> 08 >> 32), y el nombre real del fichero será 82628192407436799614057093114020. Como se puede notar, en el caso que un usuario deposite un ODE, que sea por ejemplo el archivo *xxx.pdf* desaparece toda referencia al mismo como tal y es sustituida por este tipo de referencia interna usando bitstream (Figura 4).



Figura 4: Estructura de directorios donde se almacena la información de los ODE

Los metadatos, información técnica sobre los datos, se almacenan junto a los documentos para apoyar la preservación. Si bien los ítems están almacenados en la base de datos, las consultas que los usuarios ingresen al buscador no extraerán la información directamente de los archivos sino que la obtendrán a partir de los metadatos asociados. De esta manera, es indispensable que se le proporcione a DSpace la mayor cantidad de metadatos y que sean lo más precisos posibles ya que de ello depende el éxito de las búsquedas. Cuanto más claros, extensos y accesibles sean los metadatos que se den acerca de los ODE, más fácil será encontrarlos y utilizar la información.

Finalizada la etapa de instalación, se configuró DSpace, tanto en apariencia como en funcionamiento. Con respecto a la apariencia se adaptó la interfaz de la página principal de DSpace mediante el uso de logos, colores institucionales y una breve descripción de la finalidad del repositorio.

3.2. Estándares, metadatos y extracción

Como se conoce, existen gran cantidad de estándares de metadatos para describir ODEs. Dada la importancia que tienen los metadatos en este tipo de sistemas, fue necesario hacer un estudio detallado de dichos estándares para definir su uso en la implementación particular que se está llevando a cabo.

En trabajos previos se han analizado los distintos estándares de metadatos usados para la descripción de ODE [5,9]. Si bien DC es uno de los estándares más ampliamente utilizados, se identificó que no eran suficientes para representar las características particulares de un ODE, dado que se omitían descriptores del tipo educativo o pedagógico que eran necesarios considerar. Dichos metadatos son contemplados por el estándar LOM en su categoría *Educational*. Por lo tanto se tomó la decisión de utilizar ambos estándares en el RI en cuestión. Se realizó una clasificación de los ODE en colecciones. A partir de estas colecciones, se identificó que se necesitaban distintos metadatos para su correcta y completa descripción. Así, se propusieron diferentes metadatos para cada una de ellas. Se consideraron las siguientes colecciones para el repositorio en cuestión [10]: Libros; Capítulo de libros; Revistas; Artículo en Revista con referato; Comunicaciones en Congresos, Jornadas, Conferencias/Póster; Tesis de Posgrado: Maestría / Doctorado; Tesis de Grado: Proyectos finales de carrera / Tesinas / Tesis; Producción en Investigación: Proyecto de Investigación / Informe de Avance / Informes Técnicos / Documento de Trabajo / Software / Póster; Producción en tecnología: Patente / marca / modelo industrial / modelo de utilidad / Informe técnico / Documento de trabajo / Software / Póster; Producción en extensión: Convenios / Actas Acuerdo / Informes Técnicos / Documento de Trabajo / Software; Producción en arte: Película / documental / Videgrabación / Fotografía / Sonido / Música / Software / Póster; Producción de material didáctico; Reseña: Revisión de artículo / Revisión literaria; Datos primarios: Conjunto de datos / plano / Mapa / Imagen Satelital / Radiografía; Otros.

Cada una de estas colecciones fue descrita a través de un conjunto de metadatos que se consideraron específicos y adecuados, permitiendo su localización en forma económica y eficiente. Para cada una de estas colecciones se definió una tabla de metadatos obligatorios y optativos y se determinó el valor que estos

podrían adoptar siguiendo las directrices del SNRD. Por ejemplo, para la colección *Libro* el metadato *dc:creator* es obligatorio pero para la colección *Revista* no lo es. También se identificaron metadatos equivalentes entre DC y LOM con el objetivo de permitir búsquedas que identifiquen cualquiera de estos metadatos. Por ejemplo, se identificó la equivalencia entre *dc:creator* y *lom:lifecycle.contributor.role.author* como también entre *dc:subject* y *lom:general.keyword*.

3.3. Algoritmo de extracción de metadatos

Luego de haber definido las colecciones y los metadatos necesarios para describirlas, se desarrolló un algoritmo de extracción semiautomática de metadatos el cual se muestra en la figura 5. Este algoritmo combina alguna técnicas ya probadas tales como Alchemy [11], ParsCit [12] y uso de plantillas XML [13] con metadatos, proporcionando mejores resultados. En una primera implementación este algoritmo soportará tres formatos de archivos de entrada con extensiones: .doc, .ppt y pdf. Sin embargo el algoritmo está diseñado para ser extensible y soportar la incorporación de otros formatos sin necesidad de rediseñar el mismo.

El proceso de extracción inicia cuando *el usuario selecciona la colección* bajo la cual archivará el ODE en el Repositorio. Luego, el usuario carga uno o más archivos que desea almacenar en el repositorio. Inicialmente, el algoritmo será implementado para trabajar con archivos en formato PDF, PPT y Word.

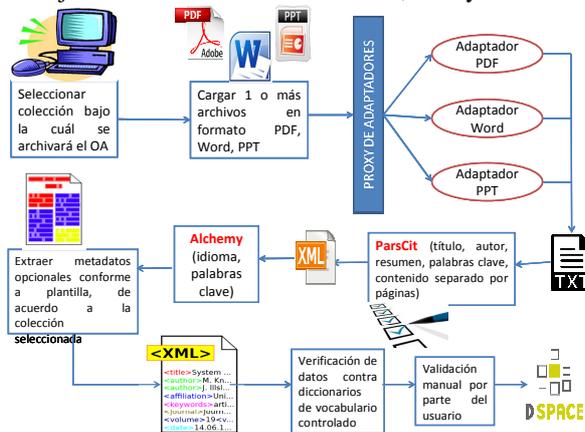


Figura 5: Algoritmo de extracción de metadatos

Estos archivos serán enviados al módulo *proxy de adaptadores*, quien se encarga de determinar el formato y enviar el archivo a su correspondiente *adaptador*. Se define un módulo *adaptador* por cada tipo de formato reconocido por el algoritmo, en este caso se definieron los módulos *Adaptador PDF*, *Adaptador Word* y *Adaptador PPT*. Estos adaptadores trabajan

transformando el archivo de entrada en un archivo con formato estándar TXT para de esta manera poder continuar el proceso independiente del formato original. Este proceso también libera de palabras *Stop Words*². Posteriormente, el texto que se extrae se pasa al *analizador de estructura ParsCit* [12], el cual genera un archivo XML donde se pueden identificar la mayoría de los metadatos descriptivos obligatorios definidos en [10], además del contenido separado por páginas del archivo original. El archivo XML generado se envía a *Alchemy* [11], para extraer los metadatos adicionales de idioma y palabras clave.

Se utiliza la combinación de ParsCit + Alchemy, teniendo en cuenta que de acuerdo a lo presentado en [14] se logró incrementar la calidad de las palabras claves retornadas pasando de un 56% de resultados correctos obtenidos solo con Alchemy, a un 70% de precisión.

El siguiente paso en el proceso es extraer algunos metadatos opcionales con apoyo de una *plantilla*, aplicada conforme a la colección seleccionada por el usuario en el primer paso del proceso de extracción. La figura 6 muestra un ejemplo de plantilla para la categoría *Libro*. Como se observa, las distintas etiquetas del archivo corresponden a metadatos optativos de dicha categoría. Todos los metadatos extraídos se consolidan en un archivo XML a partir del cual se limpian, normalizan y envían a un *módulo de verificación contra diccionarios de vocabulario controlado*, bajo el concepto de distancia de edición, donde se elige la palabra con la mínima distancia entre el metadato extraído, conforme al Algoritmo de Distancia Levenshtein [15].

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<rights></rights>
<holder></holder>
<fil></fil>
<rights_text></rights_text>
<editor></editor>
<illustrator></illustrator>
<contributor_other></contributor_other>
<publisher></publisher>
<issued></issued>
<identifier></identifier>
<publisher_version></publisher_version>
<ispartof_series></ispartof_series>
<source></source>
<sponsorship></sponsorship>
```

Figura 6: Plantilla usada para la extracción de metadatos optativos de la colección Libro

² Palabras comunes, conocidas como *palabras vacías*, que típicamente modifican otras palabras, pero que no acarrear un significado inherente en sí mismas, tales como adverbios, conjunciones o preposiciones.

Por último, los metadatos validados se consolidan en un archivo XML y se presentan al usuario para que realice la respectiva *validación manual* para corregir y completar la información, en especial de los metadatos descriptivos que sirven para la representación e identificación de los recursos y admiten la búsqueda y recuperación de la información, con el fin de enviarlos a DSpace para finalizar el depósito del ODE en el repositorio.

3.4. Definición de una ontología DC para la interoperabilidad semántica del RI

Como se mencionó anteriormente, surge la necesidad de estandarizar la manera de describir los objetos almacenados, de forma tal que se permitan búsquedas más concretas de los recursos disponibles, como así también la reutilización de los mismos. Para lograr este objetivo, la utilización de DC o LOM es insuficiente por lo que se propone una ontología denominada *DCOntoRep* [16], que dé soporte a la descripción de ODE de una manera estandarizada, considerando no sólo los estándares DC y LOM sino también incorporando las directivas de SNRD a través de la definición de axiomas, reglas de derivación e instancias.

Una ontología es una especificación formal y explícita de una conceptualización compartida [17]. Por lo tanto debe tener una especificación, una definición y debe ser explícita teniendo en cuenta el contexto en el que se la está utilizando; además de que la misma es compartida, por lo cual debe ser entendida por al menos dos actores involucrados en la especificación.

La ontología que a continuación se presenta, se desarrolló siguiendo los pasos sugeridos por el método basado en Meta-Modelo de Ingeniería de Procesos de Software y Sistemas, versión 2.0, Software & System Process Engineering Meta-Model, SPEM.0 [18]. En su desarrollo, se tuvieron en cuenta las directrices SNRD y se formularon algunas preguntas de competencia que permitan la validación. Posteriormente, se analizaron las directivas del SNRD para enriquecer la ontología agregando relaciones, axiomas y reglas que fueran necesarias.

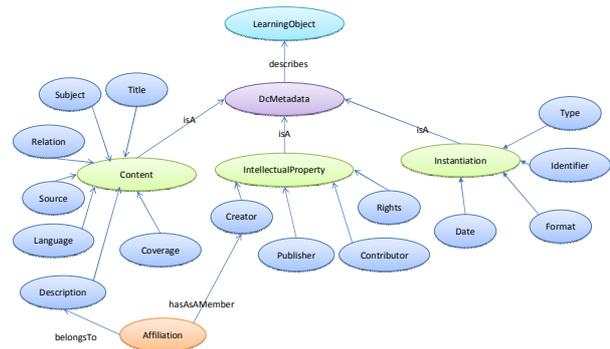


Figura 7: Ontología de metadatos

El gráfico de la figura 7 muestra los principales conceptos definidos en la ontología *DCOntoRep*. Se define el concepto *LearningObject* para representar los objetos de aprendizaje (LO por su sigla en inglés, en este contexto ODE y LO son considerados equivalentes) que son descritos a través de metadatos. El concepto *DcMetadata* abstrae los metadatos de DC. Como subclases de este concepto se definieron tres clases correspondientes a cada una de las categorías especificadas en DC: (i) *Content*, que agrupa los metadatos para describir el contenido de un recurso, (ii) *IntellectualProperty* que agrupa los metadatos que describen las cuestiones referentes a la propiedad intelectual y derechos de uso del recurso y finalmente (iii) *Instantiation* correspondiente a los metadatos que describen cuestiones técnicas de la instancia del recurso que describen.

Como subclases de *content* se definieron los conceptos: *Title*, para contener el título del LO; *Subject*, que representa el dominio del LO; *Description* permite describir el LO a través de un texto significativo; *Source*, identifica un objeto digital que se ha tenido en cuenta para la formulación del LO; *Language* identifica el lenguaje correspondiente al LO; *Relation* permite identificar otras versiones del LO que se describe; *Coverage* se refiere al alcance o ámbito del LO, generalmente incluye la ubicación espacial, un período de tiempo o la jurisdicción. Para el concepto *IntellectualProperty* se representaron las siguientes subclases: *Contributor* que identifica el/los colaborador/es que hayan aportado al contenido del LO, el cual puede ser una entidad o una persona con cargo como por ejemplo directores, supervisores, editores, entre otros; *Creator* que identifica el/los autor/es principal del LO; también puede designar una institución o un evento; *Publisher* corresponde al publicador o editor que hace posible que el recurso esté disponible (puede ser una persona, una organización o un servicio); *Right* identifica los derechos de autor asociados al LO. Luego como subclases de *Instantiation* se identificaron los siguientes conceptos: *Date* que especifica la fecha de creación o

disponibilidad del LO; *Type* que establece el tipo de resultado científico, como puede ser un artículo o un libro entre otros; *Format* que identifica el formato físico o digital del LO; e *Identifier* hace referencia a una URL o URI que identifica unívocamente al LO.

4. Conclusiones

En este trabajo se presentaron los distintos aspectos abordados en la implementación de un RI para una universidad pública. Se tuvieron en cuenta las características particulares de este tipo de institución y del material que se espera sea depositado. Para tal fin se analizaron las colecciones y los metadatos para cada uno de ellos. También se trabajó sobre un extractor de metadatos que ayude al usuario en la tarea muy tediosa de completar los valores correctamente. Asimismo, se consideró necesario no restringir el conjunto de metadatos a un estándar sino proponer combinaciones de manera que el ODE quede correctamente identificado. Dado que se trata de una Universidad, la categoría *educational* de LOM debía ser tenida en cuenta. Es por eso que no sólo se incorporaron otros metadatos a los de DC, sino que también se trabajó en encontrar equivalencias entre estos. Por último se trabajó sobre el desarrollo de una ontología DC la cual modela los metadatos de DC y las recomendaciones del SNRD, incorporando aspectos semánticos en el repositorio. De esta forma, no sólo se puede contar con mecanismos de inferencia entre los metadatos sino que también esta ontología, al ser implementada en el repositorio, proveerá una forma de restringir los valores de los metadatos de acuerdo a estándares tales como ISO 639 e ISO 3166 ya que fueron importadas ontologías que representan estos estándares.

Actualmente, el repositorio se encuentra en la etapa de prototipo y está siendo probado en un centro de investigación para depurar errores e introducir mejoras. Cada uno de los aspectos que se discutieron en este trabajo, está siendo desarrollado por diferentes grupos. El algoritmo de extracción automática de metadatos está en etapa de implementación, se muestra en este trabajo el diseño del algoritmo.

Como trabajos futuros se espera completar el extractor de metadatos y la implementación de la ontología. También se trabajará sobre los motores de búsqueda que propone Dspace de manera de mejorarla en base a la ontología formulada como así también incorporar en estas búsquedas los metadatos de LOM que fueron considerados en este repositorio.

5. Referencias

- [1] AENOR, 2009. Asociación Española de Normalización y Certificación. www.aenor.es. UNE 71361. Perfil de aplicación LOM-ES para etiquetado normalizado de Objetos Digitales Educativos (ODE).
- [2] San Martín, P., Guarnieri, G., & Bongiovani, P. Propuesta sociotecnológica para el desarrollo de repositorios de Acceso Abierto adecuados al contexto universitario argentino. *E-Ciencias De La Información*, 4(2), 2014. [Disponible en: <http://dx.doi.org/10.15517/eci.v4i2.15131> Consultado: Agosto 2015]
- [3] National information standards organization, "The Dublin core metadata element set", ISSN 1041-5635, 2013
- [4] IEEE Standard for Learning Object Metadata. Learning Technology Standards Committee of the IEEE Computer Society. 1484.12.1-2002
- [5] Sandobal, V.; Ale, M.; Gutiérrez, M.: Hacia la integración de la producción científica y pedagógica de las universidades nacionales a los Repositorios de Acceso Abierto - Un análisis situacional. III Jornadas de investigación en ingeniería del NEA y países limítrofes. Tecnologías de Información y comunicaciones. ISBN: 978-950-42-0157-1. pp 46-53. Junio 2014, e-book <http://jiim.frre.utn.edu.ar/actas/2TIC/index.pdf>
- [6] SNRD. Directrices SNRD: Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales Ministerio de Ciencia, Tecnología e Innovación Productiva, 2013.
- [7] Mina, A., Finós, F., Pérez, M., Cifre, S.: R.I.U.T.N: Prototipo de Repositorio Institucional de acceso abierto para la UTN FRSF. CONAIIISI 2014 Simposio de trabajos estudiantiles. San Luis 2014.
- [8] COAR: Confederation of Open Access Repositories. The Current State of Open Access Repository Interoperability. Working Group 2: Repository Interoperability. Octubre 2012.
- [9] Pinilla, Adriana. Extracción automática de metadatos a partir de objetos de aprendizaje en un repositorio institucional: Estado del arte. Trabajo Final integrador Especialización ISI. UTN – FRSF, 2014.
- [10] Giorgetti, C., Romero, L., Gutiérrez, M. Definición de Metadatos Educativos para Repositorios de Instituciones de Educación Superior. 2015.
- [11] Alchemy API. 2014. [Disponible en: <http://www.alchemyapi.com> Consultado: Octubre 2014]
- [12] ParsCit: An open-source CRF Reference String and Logical Document Structure Parsing Package. 2014. [Disponible en: <http://wing.comp.nus.edu.sg/parsCit/index.html> Consultado: Agosto 2015]
- [13] Flynn, P., Zhou, L., Maly, K., Zeil, S. & Zubair., M. (2007). Automated template-based metadata extraction architecture, Proceedings of the 10th international conference on Asian digital libraries: looking back 10 years and forging

- [14] Casali, A., Deco, C., Fontanarrosa, S. & Sabater, C. Asistente para el depósito de objetos en repositorios con extracción automática de metadatos. XV Simposio Internacional de Tecnologías de la Información y las Comunicaciones en la Educación (SINTICE 2013). Madrid, España, 2013. ISBN: 978-84-695-8362-3. new frontiers, December 10-13, 2007, Hanoi, Vietnam
- [15] Levenshtein, V. Binary codes capable of correcting deletions, insertions and reversals. Doklady Akademii Nauk SSSR, 163(4), p. 845-848, 1965
- [16] Sandobal, V., Ale, M., Gutiérrez, M. DCOntoRep: hacia la interoperabilidad semántica de Repositorios Institucionales de Acceso abierto. Proceeding JAIIO – SAOA p. 91-100, 2015
- [17] Studer R., Benjamins VR, Fensel D. Knowledge Engineering: Principles and Methods. IEEE Transaction and Data and Knowledge Engineering 25(1-2) :161-197. 1998
- [18] Software & System Process Engineering Metamodel Specification (SPEM). Version 2.0 Object Management Group, 2008. www.omg.org/spec/SPEM/2.0/