

Primeros pasos hacia la Preservación de Objetos de Aprendizaje: una comparación entre el estándar de metadato utilizado en Dspace y el diccionario de datos PREMIS

Ing. Valeria Sandobal Verón¹; Ing. María Alejandra Cernadas², Mag. Liliana Cuenca Pletsch³;
Universidad Tecnológica Nacional, Facultad Regional Resistencia
French 414 – Resistencia (3500) - Chaco

1. Jefe de Trabajos Prácticos. vsandobal@fre.utn.edu.ar – 0362-154 664580
2. Jefe de Trabajos Prácticos. ma_cemadas@hotmail.com – 0362-154 302755
3. Prof. Titular. cplr@fre.utn.edu.ar – 0362-154563171

Resumen

A continuación se presenta una propuesta de incorporación de etiqueta de metadatos a la estructura Dublin Core utilizada en el repositorio Dspace de la Facultad Regional Resistencia, Universidad Tecnológica Nacional, con el objetivo de aproximarnos a través de metadatos a técnicas de preservación de objetos digitales. Esta propuesta se basa en analizar el diccionario de datos PREMIS y los metadatos de DC, mapeando estos últimos o agregando las etiquetas que se consideren necesarias. El análisis de la incorporación o mapeo de los metadatos propuestos se basa en que un metadato de preservación debe: guardar información técnica que dé soporte a las decisiones y acciones de preservación, documentar las acciones de preservación que se encuentran planificadas en el repositorio, resguardar los efectos de las estrategias de preservación, garantizar la autenticidad de los objetos digitales y la gestión de derechos.

Palabras Clave: metadatos, preservación, OAIS, PREMIS

Introducción

La preservación constituye un aspecto fundamental en el tratamiento de imágenes y objetos digitales y requiere de actividades tales como: copias de seguridad, reorganización de ficheros, visionado, verificación, etc. Para poder realizar de forma efectiva estas actividades resulta necesario planificar la manera en la que se llevarán a cabo las mismas. En este sentido, es recomendable el desarrollo de estrategias y pautas de actuación que tengan en cuenta las normas y directrices nacionales e internacionales; así como también la utilización de formatos de archivo y sistemas de almacenamiento estándar para asegurar la vida de la colección. La definición de estas estrategias permitirán que las colecciones puedan consultarse a corto y largo plazo. [1]

Según Conway [3] en su libro “La preservación del mundo digital”, a medida que se va experimentando con las capacidades tecnológicas, tanto los bibliotecarios como los archivistas han llegado a la definir la preservación de varias maneras. Todas las definiciones hacen referencia a garantizar la disponibilidad del Objeto Digital (OD) en el tiempo, independientemente de la tecnología con que el OD fue creado y la disponible en cada momento. La digitalización de documentos valiosos resultó en una buena forma de preservación de los objetos originales, teniendo en cuenta que al crear una copia de alta calidad de un ejemplar original, permite el acceso al objeto digitalizado y no resulta necesario el acceso físico al documento original. La preservación digital está centrada en la elección de los medios de almacenamiento, la expectativa de vida de un sistema y la posibilidad de pasar los archivos a sistemas futuros.

Para poder llegar a asegurar la disponibilidad a largo plazo es necesario establecer estrategias de preservación que pueden clasificarse en: técnicas, de organización y gestión.

En relación con las **estrategias técnicas**, están involucradas las siguientes actividades:

Mantenimiento: Se considera una actividad continua dentro de la preservación de objetos digitales y comprende el almacenamiento de las imágenes y de los archivos en medios seguros y fiables, y la realización periódica de copias de seguridad. También se encuentra relacionado con garantizar el buen estado de los objetos digitales.

Actualización: Se considera una parte esencial de la política de preservación a largo plazo, sin embargo no se considera una estrategia de preservación en sentido pleno. Se pretende que el formato que soporta el objeto digital no quede obsoleto, por lo cual debe actualizarse a herramientas actuales.

Migración: Proceso de transferencia de información digital de una configuración a otra o de una generación de ordenadores a otra. Teniendo en cuenta, que durante el proceso de migración, algunas propiedades importantes del objeto pueden perderse, es necesario que la información que se obtiene al momento de incorporarlo al Repositorio Institucional sea lo más específico posible.

Emulación: Comprende la recreación del entorno técnico requerido para ver y utilizar la colección digital. Para lo cual resulta necesario conocer los requisitos técnicos del equipo, como así también el software que debe instalarse.

Preservación de la tecnología: Consiste en preservar el entorno técnico que hace funcionar el sistema como son sistemas operativos, programas y aplicaciones originales, entre otros.

Arqueología digital: Incluye métodos y procedimientos para rescatar contenidos dañados de programas o entornos obsoletos o afectados por cualquier anomalía

En cuanto a las **estrategias de organización** son las que acompañan a las estrategias técnicas y están relacionadas con: la contratación y capacitación de personal, entre otros. Además debe estar dirigido por la misión y objetivos de la institución, en relación a hacer explícito qué materiales deben ser conservados y por qué; también deben establecerse las normas y criterios a seguir. Resulta conveniente también que las instituciones desarrollen un plan de actuación y que el mismo sea ejecutado de forma tal de evaluar los requisitos de preservación, obteniendo de esta manera el acceso a largo plazo, como así también identificar los costos asociados a esta actividad y valorar los riesgos de su implementación. Estos aspectos también forman parte de las **estrategias de gestión**.

Metadatos y Metadatos de Preservación

Los metadatos se utilizan para la descripción del objeto digital, permitiendo de esta manera facilitar la búsqueda y localización de los objetos digitales almacenados en la red. La norma ISO15489-1 del 2001 define los metadatos, en el contexto de la gestión de documentos, como: "datos que describen el contexto, contenido y estructura de los documentos, así como su gestión a lo largo del tiempo(...) Como tales, los metadatos son información estructurada o semiestructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo y dentro de un mismo dominio o dominios diferentes."

Los mismos pueden ser categorizados de la siguiente manera:

Metadatos descriptivos: permite la localización y recuperación de objetos digitales de una determinada categoría de documentos o imágenes.

Metadatos estructurales: proporcionan información sobre la estructura interna de los objetos digitales como ser: página, sección, índice, tabla de contenidos, entre otros. También permiten la descripción de las relaciones entre los objetos digitales como por ejemplo: la fotografía A está incluida en el objeto digital B)

Metadatos administrativos: incluye datos técnicos como la creación y el control de calidad, gestión de derecho, control de acceso y utilización y condiciones de preservación (ciclos de actualización, migración, etc)

Como lo menciona Méndez Rodríguez [4] "Los metadatos son el fundamento de los depósitos digitales ya que hacen útiles e interoperables los datos transcribiendo la información necesaria si se aplican con principios de calidad. Deben ser precisos y no demasiado específicos y apoyarse en un entorno informático que incluya metadatos administrativos, estructurales y descriptivos."

Según Senso y Piñeiro [7] la importancia de los metadatos se basan en las siguientes razones: *Incrementan la accesibilidad* (si se realiza la descripción de objetos digitales mediante un conjunto de metadatos, esto aumenta la posibilidad de que el objeto digital pueda ser accedido y permite la

búsqueda de objetos digitales en diferentes colecciones); *Disminuye del tráfico en la red* (al indizar la representación del objeto, y no el objeto en sí, no requiere demasiado ancho de banda para hacer las búsquedas); facilita la difusión de versiones digitales de un único objeto; *Control de versiones* (se refiere tanto a la definición a partir de metadatos del ciclo de vida de un objeto, como así también a generar distintas cantidades de información sobre un mismo objeto con el fin de distribuirla a un público heterogéneo); *Aspectos legales* (a través de los metadatos es posible restringir el uso y acceso al OD descrito incluyendo información como: derechos de autor, control del uso de todo o una parte del objeto, como así también el resguardo de la autoría) y la *Preservación del objeto original*.

En particular y para los fines de este artículo nos concentraremos en los metadatos de preservación (recordkeeping metadata), que no pueden considerarse como una categoría separada de metadatos, sino la suma de varios tipos. Estos metadatos pueden incluir elementos de carácter único y/o con mayor nivel de detalle que los metadatos empleados para otras funciones. Acordando con [6] “entenderemos por metadatos de preservación como la información necesaria para mantener la viabilidad, la representación y la comprensión de recursos digitales a largo plazo. La *viabilidad* requiere que la cadena de bits de un objeto digital archivado permanezca intacta y legible a partir de los medios digitales sobre los que se ha almacenado. La *representación* se refiere a la traducción de la cadena de bits a una forma que pueda ser vista por usuarios humanos, o procesada por ordenadores. La *comprensión* implica proporcionar suficiente información para que el contenido representado pueda ser interpretado y comprendido por sus usuarios pretendidos. Los metadatos de preservación pueden servir como entrada a los procesos de conservación, y también registrar la salida de estos mismos procesos”

Materiales y Métodos

Modelo de referencia OAIS

El foco de la preservación está en mantener la información de los objetos digitales asegurándose que se encuentran en un medio accesible a largo plazo y el medio en el que se encuentra actualmente pueda ser controlado. El gran crecimiento de la información digital, en cuanto a cantidad como así también en los formatos en que se encuentran, acompañado del avance tecnológico en cuanto a hardware y software hace que los medios utilizados queden rápidamente obsoletos. Por lo cual urge la necesidad de establecer estrategias que permitan la preservación de la información digital disponible actualmente.

El modelo de referencia OAIS, un estándar de la ISO – ISO 14721:2003 propone un marco de referencia para la preservación de la información de colecciones digitales y el acceso a largo plazo de los mismos. En este sentido, además presenta un nuevo enfoque sobre la función de los metadatos en la preservación.

El entorno de OAIS está formado por: *Productor* (personas, sistemas clientes que proveen la información a ser preservada); *Gestor/Administrador* (entidad responsable de establecer las políticas generales que deberán llevarse a cabo para el archivo, tales como la selección y la financiación. No es el encargado de la administración del día a día del OD); *Consumidores* (son los que utilizan la información preservada. Una clase especial de consumidores es la “Designated Community” (Comunidad designada) que representa un subconjunto de consumidores que tiene la capacidad para entender la información preservada).

Para que el objeto de datos sea correctamente interpretado, a este entorno se le agregan dos componentes: la base de conocimiento del usuario que contiene un conjunto de referencias que le permite interpretar información de acuerdo al contexto en el que se desarrolla; y la información sobre la representación, que es información externa a la base de conocimiento del usuario, cuyo objetivo es llegar a una interpretación correcta.

En el entorno OAIS un objeto de información, contiene información estructural e información sobre contenido, y está compuesto por el objeto de datos y la representación de la información para interpretarlo. Ambos elementos constituyen lo que se denomina “objeto de información”, el cual

contiene información estructural e información sobre el contenido. A la vez, esta información de contenido está compuesta por objetos de datos de contenido y representación de la información asociada. Si a la información de contenido le agregamos la información sobre las actividades de preservación, obtenemos lo que se denomina “Paquete de información”, el cual está compuesto por cuatro objetos de información, a saber: información sobre el contenido, sobre la descripción de la preservación, de empaquetamiento e información descriptiva.

El modelo de referencia OAIS establece cuatro categorías de metadatos, como información sobre preservación: *Referencia* (identificadores o códigos que asignan de forma unívoca al archivo, tanto a nivel interno como externo, ejemplos de estas referencias son ISBN, URN, entre otros); *Contexto* (relaciones entre el objeto digital y el entorno, tales como: por qué fue creado, relaciones con otra documentación); *Procedencia* (origen del objeto digital, incluye información como por ejemplo: orígenes, intervenciones para garantizar la preservación, integridad, entre otros); *Autenticación* de la información (mecanismos que se deben implementar para asegurar que la información contenida en el OD no ha sido alterada, como ser checksum o firma digital). Además de las categorías antes mencionadas el modelo OAIS agrega una quinta, denominada Representación de la información, que contiene información vinculada con los programas necesarios para procesar e interpretar de manera correcta el objeto digital.

PREMIS

El grupo PREMIS (PREservation Metadata: Implementation Strategies) surge en el 2003 con el objetivo de “*desarrollar un conjunto fundamental de metadatos de preservación, basados en directrices y recomendaciones de creación, gestión y uso, que pudiesen ponerse en práctica y aplicarse de manera general en una amplia variedad de contextos de preservación digital*”[8]. Su trabajo concluyó dos años más tarde “*con la publicación del documento Diccionario de Datos para Metadatos de Preservación: Informe final del grupo de trabajo PREMIS.*” El documento contiene un Diccionario de Datos propio, material práctico para la puesta en práctica de metadatos de preservación, y un esquema XML que permite incorporar el Diccionario de Datos en sistemas de gestión de objetos digitales. PREMIS se utiliza fundamentalmente en la etapa de diseño de un repositorio, para su evaluación y para el intercambio de los paquetes de la información archivada entre repositorios de preservación. También pueden utilizarlo como guía quienes diseñan software para repositorios. En los casos de repositorios de preservación, PREMIS es útil como un checklist. También es útil en los casos de exportación de la información almacenada para la ingesta en otro repositorio.

Metadatos fundamentales de preservación

El Diccionario de Datos PREMIS define los metadatos de preservación como *la información que un repositorio utiliza para llevar a cabo el proceso de preservación digital*. Estos metadatos engloban una serie de categorías de metadatos que normalmente se usan para diferenciarlos: **administrativos** (incluidos derechos y permisos), **técnicos** y **estructurales**. Se prestó especial atención a la documentación sobre la procedencia digital (la historia de los objetos) y a la documentación de las relaciones, especialmente aquellas entre distintos objetos dentro del repositorio de preservación.

El Diccionario de Datos debe ser totalmente independiente de la forma en que se aplicará. Es decir, los metadatos fundamentales definen la información que un repositorio necesita, y esto es independiente del modo en que se almacena esa información, e incluso del hecho de que pueda no llegar a almacenarse. El Diccionario define un conjunto de unidades semánticas, propiedades e información que la mayoría de los repositorios necesita conocer de sus entidades para asegurar la preservación.

El modelo de datos PREMIS

EL modelo de datos define entidades para organizar las unidades semánticas. Las cinco entidades consideradas importantes son: Entidades Intelectuales, Objetos, Acontecimientos, Derechos y Agentes.

Cada unidad semántica definida en el Diccionario de Datos es una propiedad de una de las entidades del modelo de datos.

Las entidades en el modelo de datos PREMIS se definen de la siguiente manera:

Entidad Intelectual: conjunto de contenidos que se considera una única unidad intelectual a efectos de gestión y descripción, por ejemplo, un libro, un mapa, una fotografía o una base de datos. Una Entidad Intelectual puede comprender otras Entidades Intelectuales y puede tener una o más representaciones digitales; **Objeto [digital]:** unidad discreta de información en formato digital. Se clasifican en tres tipos: **archivo, representación y cadena de bits.** El *archivo* puede ser un PDF un .doc, el objeto *representación* es el conjunto de todos los archivos (un libro) y los de *cadena de bits* son subconjuntos de archivos con propiedades útiles a la preservación, por ejemplo la tapa puede tener sus propios identificadores; **Acontecimiento:** acción que al menos afecta a un Objeto o Agente asociado o conocido por el repositorio de preservación. También se lo define como “Evento”; **Agente:** pueden ser personas, organización o aplicaciones informáticas asociadas a los Acontecimientos durante la vida de un Objeto, o a los Derechos ligados a un objeto; **Derechos:** declaración de uno o varios derechos o permisos pertenecientes a un Objeto o Agente. En un repositorio puede quedar registrada una gran variedad de información sobre los derechos, sin embargo, el corpus mínimo de información de derechos que un repositorio de preservación debería conocer es el que hace referencia a los derechos o permisos con los que un repositorio tiene que cumplir y que corresponde a las acciones que afectan a los objetos dentro del mismo. Esta información puede quedar garantizada por la ley de copyright, por una norma, o por un acuerdo de licencia con el titular de los derechos.

El Diccionario de Datos PREMIS define **unidades semánticas**. Cada unidad semántica especificada en el Diccionario de Datos se mapea a una de las entidades del modelo de datos. Una unidad semántica puede entenderse como una propiedad de una entidad. Por ejemplo, la unidad semántica *size* es una propiedad de la entidad Objeto. Además, los Acontecimientos tienen consecuencias. Por ejemplo, un acontecimiento de migración crea un archivo que ha perdido alguna característica, la pérdida de la característica puede considerarse una consecuencia del Acontecimiento y, por lo tanto, una propiedad de la entidad Acontecimiento.

En algunos casos, una unidad semántica adquiere la forma de un **contenedor** que agrupa un conjunto de unidades semánticas relacionadas. Por ejemplo, la unidad semántica *identifier* (identificador) agrupa las unidades semánticas *identifierType* (tipo de identificador) e *identifierValue* (valor del identificador). Las subunidades agrupadas se conocen como los **componentes semánticos** del contenedor. Algunos contenedores se definen como **contenedores de extensión**, que permiten el uso de metadatos codificados según un esquema externo. Esto posibilita la extensión de PREMIS con elementos de metadatos de detalle, es decir, elementos que no se consideran fundamentales o que están fuera del ámbito del Diccionario de Datos.

Una **relación** es una declaración de asociación entre casos individuales de entidades. Por ejemplo, la afirmación «el Objeto A tiene un formato B» puede considerarse una relación entre A y B. El modelo PREMIS, sin embargo, trata el formato B como una propiedad del Objeto A. PREMIS reserva el concepto de relación para asociaciones entre dos o más entidades Objeto o entre entidades de distintos tipos, tales como un Objeto y un Agente.

El Repositorio de la Facultad Regional Resistencia

En la Facultad Regional Resistencia de la Universidad Tecnológica Nacional (UTN-FRRe), en el marco del proyecto “Modelización de un repositorio de objetos de aprendizaje para la gestión del conocimiento para la UTN-FRRe”, se ha decidido la implementación de Dspace ya que el mismo es de código abierto y se adapta a los objetivos del proyecto. Este software permite contar con una organización para capturar y describir material digital, como así también la posibilidad de realizar búsquedas, almacenar y preservar objetos digitales a largo plazo. Dspace provee una estructura de 5

componentes para almacenar la información: *Comunidades/Subcomunidades* (pueden incluir metadatos que describan un área, departamento, etc); *Colecciones* (agrupaciones de ítems formados por metadatos y archivos); *Bundles* (paquetes de archivos) y *Bitstreams* (archivos almacenados). Estos componentes, y en particular las comunidades y colecciones, pretenden representar la organización que está utilizando el repositorio Dspace. Así por ejemplo, para la UTN-FRRe, podríamos definir como Comunidad la Secretaría de Ciencia y Tecnología; una subcomunidad estaría definida por los Eventos que organiza esta Secretaría y las colecciones estarían dadas por libros, actas de congresos y resoluciones entre otros.

Para una correcta administración de los ítems que se almacenarán en Dspace, es posible definir flujos de trabajo que serán controlados por diferentes usuarios a medida que el ítem va avanzando en la aprobación hasta llegar a ser publicado. Las opciones de flujo de trabajo que se pueden definir para el envío y publicación de un ítem se detallan a continuación:

Aceptar/Rechazar: al seleccionar esta opción se genera un paso en donde antes que el ítem sea publicado deberá pasar por un control de aceptación o rechazo por los usuarios habilitados para realizar esta acción; sin la posibilidad de modificar los metadatos. Este primer control es conveniente cuando se pretende llevar un control estricto de los ítems que serán publicados en las colecciones. Al ser aceptado este ítem continúa su flujo de trabajo y pasa a la comunidad; en caso contrario se devolverá al publicador acompañado por las razones del rechazo del ítem.

Aceptar/Rechazar/Editar metadatos: al seleccionar esta opción se adicionará un paso más al mecanismo anteriormente descrito. Además de que el usuario habilitado pueda aceptar o rechazar los ítems, se agrega la opción de poder modificar los metadatos antes de finalizar el proceso de publicación del ítem. Al igual que el flujo de trabajo anterior se podrá aceptar, en cuyo caso pasará a estar publicado en la comunidad o rechazado y deberá comunicarse al publicador sobre esta decisión y los motivos de la misma.

Editar metadatos: en el caso de seleccionar esta opción solo podrán editarse los metadatos que describen a un ítem, pero no podrá rechazarse para su publicación en la comunidad.

En relación con los metadatos utilizados se ha seleccionado una de las opciones predeterminadas en Dspace, tomando como base el estándar Dublin Core y agregando los cualificadores[5]:

- Alternative es un metadato que permite incluir un título alternativo al título oficial (Title) Alternative extiende al original de Title.
- Created; Valid; Available; Issued; Modified; DateAccepted; DateCopyrighted; DateSubmitted extienden la información contenida en el metadato Date (Fecha asociada con un evento en el ciclo de vida del recurso). Son usados cuando se requiere más de una fecha, y éstas pueden ser importantes para los usuarios.
- BibliographicCitation es un cualificador del metadato Identifier, que toma una referencia bibliográfica para el recurso.
- isVersionOf, hasVersion, isReplacedBy, Replaces, isRequiredBy, Requires, isPartOf, hasPart, isReferencedBy, References, isPartOf, hasFormat, ConformTo relacionados todos con el metadato de origen Relation (referencia un recurso relacionado con el OD)
- accessRights, License son cualificadores del metadato Rights (información sobre los derechos contenidos en y sobre el recurso)

Se considera que estos metadatos proporcionan una descripción más significativas de los ítems que serán publicados en las comunidades que se han creado en el repositorio Dspace.

Resultados y Discusión

A continuación se presenta una propuesta de comparación entre el esquema de metadatos que utiliza Dspace, el Dublin Core cualificado; y el diccionario de datos proporcionado por PREMIS como una

primera aproximación a la preservación de los objetos de aprendizaje que se almacenan en el repositorio de la Facultad Regional Resistencia.

A continuación se presenta la *tabla 1* que contiene las unidades semánticas del diccionario de datos PREMIS y su posible correspondencia con el estándar DC cualificado utilizado en Dspace. En la *tabla 2* se muestran las unidades semánticas/componentes semánticos de PREMIS y su correspondientes etiquetas sugeridas para ser incorporadas al esquema de metadatos.

Unidades Semánticas/Componentes Semánticos PREMIS	Etiquetas de metadatos DC cualificado DSpace
ObjectCharacteristics size format	dc.format.extent dc.format
OriginalName	dc.title
Storage contentLocation	dc.identifier.uri
Relationship relationshipType relationshipSubType relatedObjectIdentification	dc.relation.haspart, dc.relation.isbasedon, dc.relation.ispartof, dc.relation.ispartofseries, dc.relation.isversionof, dc.relation.uri, dc.relation.requires

Tabla 1: Relación Unidades Semánticas y las etiquetas presentes en el DC cuaificado de Dspace

Unidades Semánticas/Componentes Semánticos PREMIS	Etiquetas de metadatos que deberían agregarse al esquema DSpace
ObjectIdentifier objectIdentifierType objectIdentifierValue	premis.objectIdentifierType premis.objectIdentifierValue
ObjectCharacteristics - creatingApplication creatingApplicationName creatingApplicationVersion dateCreatedByApplication creatingApplicationExtension	premis.creatingApplicationName premis.creatingApplicationVersion premis.dateCreatedByApplication premis.creatingApplicationExtension
Storage - contentLocation contentLocationType contentLocationValue storageMedium	premis.contentLocationType premis.contentLocationValue premis.storageMedium
Enviroment -enviromentNote software swName swVersion swType hardware hwName hwType	premis.enviromentNote premis.swName premis.swVersion premis.swType premis.hwName premis.hwType
EventIdentifier agentIdentifierType agentIdentifierValue	premis.agentIdentifierType premis.agentIdentifierValue
EventType	premis.eventDataType

EventDateTime	premis.eventDateTime
AgentIdentifier agentIdentifierType agentIdentifierValue	premis.agentIdentifierType premis.agentIdentifierValue
RightsStatement rightsStatementIdentifier rightsStatementIdentifierType rightsStatementIdentifierValue rightsBasis	premis.rightsStatementIdentifier premis.rightsStatementIdentifierType premis.rightsStatementIdentifierValue premis.rightsBasis

Tabla 2: Unidades semánticas de PREMIS con su correspondiente etiqueta a ser agregado

El diccionario de datos de PREMIS define la unidad semántica *ObjectIdentifier* que sirve para identificar unívocamente al objeto digital dentro del sistema de preservación. Esta unidad está compuesta por los componentes semánticos: *objectIdentifierType* sirve como denominación del dominio en el que el objeto es único (debería proceder de un vocabulario controlado) y *objectIdentifierValue* es el valor de *ObjectIdentifierType*. En este caso resulta necesario agregar las etiquetas *premis.objectIdentifierType* y *premis.objectIdentifierValue*.

Otra de las unidades semánticas definidas en el diccionario es *objectCharacteristic* como aquella que hace referencia a propiedades técnicas aplicables a la mayoría de los formatos y se encuentra definida por los componentes semánticos: *compositionLevel*, *fixity*, *size*, *format*, *creatingApplication*, *inhibitors* y *objectCharacteristicsExtension*. Para nuestro análisis se han mapeado los componentes *size* y *format* con las etiquetas de metadatos *dc.format.extent* y *dc.format*. En el caso de *size* y *dc.format.extent* permiten indicar el tamaño de bytes de un fichero o cadena de bits almacenado en el repositorio. Esto permitiría desde el punto de vista de la preservación asegurar que el tamaño del documento recuperado es efectivamente el mismo que cuando se guardó. Ahora bien, para *format* y *dc.format* permite establecer el formato de un fichero o cadena de bits. En este punto es importante indicar que para mayor detalle el esquema de metadatos seleccionado cuenta con las etiquetas: *dc.format.medium*, *dc.format.mimetype*; utilizándose *dc.format* para indicar cualquier otro formato que no esté especificado en los cualificadores mencionados. Esta especificidad de los cualificadores brinda mayor detalle que resultan imprescindibles en las actividades de preservación, nos permite conocer tener más información sobre el objeto almacenado. Entre los componentes semánticos de *objectCharacteristic* se encuentra *creatingApplication* que permite definir la información relacionada con la aplicación con la que se ha creado el objeto; además este componente está compuesto por: *creatingApplicationName* utilizado para indicar el nombre del programa con la que se ha creado el objeto; *creatingApplicationVersion* determina la versión del programa y *dateCreatedByApplication* establece la fecha y hora(al menos aproximada) de creación del objeto. En este caso se considera necesaria la inclusión al esquema de las siguientes etiquetas: *premis.creatingApplicationName*, *premis.creatingApplicationVersion* y *premis.dateCreatedByApplication*.

También se incluye como unidad semántica de PREMIS *originalName* donde se guarda el nombre del objeto digital tal cual es entregado o harvesteado por el repositorio, en este caso se mapea con la etiqueta *dc.title* considerando que en la misma se guarda el nombre original del objeto y no se modifica el mismo, cualquier modificación al título deberá ser registrado en la etiqueta *dc.title.alternative*.

En cuanto a la unidad semántica *Storage* que permite almacenar información acerca de dónde y cómo se almacena un objeto digital en un repositorio cuenta con dos componentes semánticos el *contentLocation* que resguarda la información necesaria para recuperar un fichero y *storageMedium* que hacer referencia al soporte físico de almacenamiento o bien a un sistema que maneja el soporte, teniendo como objetivo definir cómo y cuándo hacer las migraciones. En el caso del componente *contentLocation* contamos con información con la etiqueta *dc.identifier.uri*; sin embargo se considera

necesario agregar además *premis.contentLocationType* y *premis.contentLocationValue* para los casos de los tipos Handle, NTF y EXT3. Para el componente storageMedium se considera que debe agregarse al esquema la etiqueta *premis.storageMedium*.

Para la unidad semántica Environment que se utiliza para guardar una combinación de información referente al hardware y software de apoyo a la utilización del objeto digital. Esta unidad tiene como componentes: environmentCharacteristic, environmentPurpose, environmentNote, dependency, software, hardware, environmentExtension. Nuestro análisis incluye a los componentes de environmentNote, software y hardware considerando que el uso de los componentes es opcional, aunque si lo usamos al menos uno de ellos debe estar presente. En el caso de *environmentNote*, que contiene información adicional sobre las características del entorno que se consideran relevantes para obtener una mayor conocimiento del mismo, se agrega la etiqueta *premis.environmentNote*. Para hacer referencia al componente software que se utiliza para indicar el software necesario para visualizar o utilizar el objeto digital se agregaron etiquetas por cada una de sus subcomponentes, a saber:

- swName se agregó *premis.swName* que se utiliza para indicar el nombre del software y empresa de software
- swVersion se agregó *premis.swVersion* que se utiliza para indicar la versión del software referenciado en swName
- swType se incluye *premis.swType* que se utiliza para indicar la clase o categoría del software, por ejemplo: renderer, ancillary, operatingSystem, driver entre otros.

Otro componente semántico de Environment, es Hardware que sirve para indicar los componentes de hardware necesarios para que el software indicado en swName funcione. este componente tiene como subcomponentes a:

- hwName, en este subcomponente se puede indicar el fabricante, modelo y versión, al esquema propuesto se agregará *premis.hwName*.
- hwType, este subcomponente se utiliza para indicar la clase o categoría del hardware, algunos valores posibles son: procesador, memoria, dispositivo entrada/salida, almacenamiento. A la estructura de metadatos se agrega *premis.hwType*.

En el caso de la unidad semántica *Relationship*, que permite describir la información de la relación existente entre este objeto digital y otro u otros objetos digitales diferentes, el diccionario de datos PREMIS propone la incorporación de los componentes semánticos *relationshipType*, *relationshipSubType*, *relatedObjectIdentification* y *relatedEventIdentification*. Para hacer referencia a estos componentes se propone utilizar las etiquetas de metadatos de DC: dc.relation.haspart, dc.relation.isbasedon, dc.relation.ispartof, dc.relation.ispartofseries, dc.relation.isversionof, dc.relation.uri y dc.relation.requires. Considerándose que contiene el detalle suficiente para conocer con que otros objetos digitales está relacionado el objeto digital descrito.

Hasta aquí se han descrito las unidades y componentes semánticos de la entidad **objeto**; como se describió en la sección anterior PREMIS propone incluir datos relacionados con lo que se denomina entidad **acontecimiento**, **agente** y **derecho**. En este sentido, como una primera aproximación se consideró agregar las etiquetas de metadatos para las unidades semánticas obligatorias para cada entidad, como se describe a continuación:

- Para el objeto acontecimiento se tuvieron en cuenta las unidades semánticas EventIdentifier, EventType y EventDateTime. En el caso *EventIdentifier* que se utiliza para identificar unívocamente el acontecimiento ocurrido dentro del repositorio. Los componentes semánticos son *eventIdentifierType* para establecer el nombre del dominio en el cual el identificador del acontecimiento es unívoco, en este caso se agrega la etiqueta *premis.eventIdentifierType*; y *eventIdentifierValue* donde se especifica el valor de eventIdentifierType, la etiqueta agregada para este caso es *premis.eventIdentifierValue*. En el caso del componente *EventType* que se utiliza para

categorizar la naturaleza de un acontecimiento, como ser: E77 - Ingest ; se incluye la etiqueta *premis.EventType*. Para el componente *EventDateTime* utilizado para indicar la fecha y la hora exactas o bien rango de las mismas durante el cual se ha producido un acontecimiento

- Para el objeto agente se analizó la unidad semántica *agentIdentifier*, cuya definición tiene por objetivo identificar unívocamente al agente del repositorio. Esta unidad está formada por los componentes *agentIdentifierType* y *agentIdentifierValue*. En el caso de *agentIdentifierType* se utiliza para indicar el nombre del dominio en el cual el identificador del agente es unívoco, como por ejemplo: LCNAF, SAN, MARC Organization Codes, URI; para este componente se agrega la etiqueta *premis.agentIdentifierType*. Para el componente *agentIdentifierValue* se agrega *premis.agentIdentifierValue* que indica el valor de *agentIdentifierType*.
- Para el objeto derechos, se analizó la incorporación de las unidades semánticas *rightsStatement* y *rightsBasis*. En el caso de *rightsStatement* que permite identificar los derechos que posee o han sido otorgados al repositorio para realizar ciertas actividades; tiene como componentes *rightsStatementIdentifier* que se utiliza para identificar unívocamente la declaración de derechos en el sistema del repositorio de preservación, para lo cual se agrega la etiqueta *premis.rightsStatementIdentifier*; *rightsStatementIdentifierType* se utiliza para designar el dominio en el cual el identificador de la declaración de derechos es único, se agrega para indicar esto la etiqueta *premis.rightsStatementIdentifierType* y *rightsStatementIdentifierValue* que es el valor concreto de *rightsStatementIdentifierType*, para la cual se agrega la etiqueta *premis.rightsStatementIdentifierValue*. Para la unidad semántica *rightsBasis* que es una designación que sirve de base para los derechos o permisos descritos en el *rightsStatementIdentifier*, donde algunos valores posibles son: copyright, licencia, dominio público, uso legítimo, entre otros; se asigna la etiqueta *premis.rightsBasis*.

Conclusiones

El trabajo aquí descrito es considerado como una primera aproximación hacia la preservación de los objetos de aprendizaje almacenados en el repositorio de la Facultad Regional Resistencia. Si bien, consideramos que se deben tener en cuenta otras cuestiones como ser las técnicas y las organizativas, creemos que es un puntapié inicial para llevar esta importante actividad dentro de un repositorio que pretende ser la cara visible de las producciones realizadas en el ámbito académico- científico-tecnológico. Esta propuesta pretende llevarse a la etapa de implementación, realizar las pruebas y los ajustes necesarios; además de avanzar en cuestiones técnicas relacionadas con la preservación, como ser la actualización, mantenimiento, emulación y migración.

Referencias

- [1] Benitez, A. S., & Rodríguez, A. Á. R. (2005). "Metadatos para la preservación de colecciones digitales". Cuadernos de documentación multimedia, 16, 21. Disponible en: http://www.anobium.es/docs/gc_fichas/doc/25AJKORYnu.pdf. Fecha de Consulta: 07/10/2015
- [2] Boté, Juanjo; Minguillón, Julià (2012). "Preservación de objetos de aprendizaje en repositorios digitales". Revista de Universidad y Sociedad del Conocimiento (RUSC). Vol. 9, N.º 1, págs. 22-35 UOC. Disponible en: <http://rusc.uoc.edu/ojs/index.php/rusc/article/view/v9n1-bote-minguillon/v9n1-bote-minguillon>. Fecha de consulta: 05/10/2015. ISSN 1698-580X
- [3] Conway, P. (2013). La preservación en el mundo digital. Publicaciones Centro Nacional de Conservación y Restauración. DIBAM, Chile.
- [4] Méndez Rodríguez, E. M. (2008). "Metadatos para la preservación digital: PREMIS". VIII Workshop Rebiun, Murcia, España, 20 y 21 de octubre.
- [5] Ferrera Fernández, Tránsito (2008). Dublin Core Cualificado. Documento de trabajo. Universidad de Salamanca. Servicio de Archivo y Bibliotecas.
- [6] OCLC/RLG Working Group on Preservation Metadata (2001). Preservation Metadata for Digital Objects: A Review of the State of the Art. Dublín: Online Computer Library Center; Mountain View: Research Libraries Group, 2001. URL: <http://www.oclc.org/research/projects/pmwg/presmeta_wp.pdf>. Fecha de consulta: 08/10/2015
- [7] Senso, J. A., & Piñero, A. D. L. R. (2003). El concepto de metadato. Algo más que descripción de recursos electrónicos. *Ciência da Informação*, 32(2), 95-106.
- [8] Sitio web de PREMIS: <http://www.loc.gov/standards/premis/>