

Análisis de Interoperabilidad en Repositorios Institucionales

Sandobal Verón, Valeria Celeste
Grupo de Investigación en Educación sobre Ingeniería (GIESIN)
Facultad Regional Resistencia - UTN
French N° 414
vsandobal@frre.utn.edu.ar

Ale, Mariel Alejandra
Gutiérrez, María de los Milagros
Centro de Investigación y Desarrollo en Sistemas de Información (CIDISI)
Facultad Regional Santa Fe – UTN
Lavaisse N° 610
{male;mmgutier}@frsf.utn.edu.ar

Abstract

El continuo avance en la implementación de repositorios institucionales, la diversidad de los sistemas actuales en relación a los formatos de metadatos y la posibilidad de que se puedan comunicar entre ellos, ha hecho de la interoperabilidad un reto. A esto hay que sumarle que al contar con diversos conceptos y diferentes niveles de interoperabilidad propuestos según diferentes autores resulta compleja la correcta implementación. Ante esta diversidad surgen diferentes áreas en las que se puede trabajar para abordar la interoperabilidad, que van desde la definición de metadatos, la comunicación entre sistemas, entre sistemas y repositorios; como así también iniciativas que buscan soluciones relacionadas con la identificación unívoca de autores y la persistencia de los objetos digitales. Si bien se considera que estas iniciativas dan solución a problemas de interoperabilidad, hay un tema transversal a todas estas: los metadatos. Para que la mayoría de los proyectos que proponen diferentes soluciones a la interoperabilidad, funcionen de manera correcta resulta necesario la definición y entendimiento común de los metadatos. Los esfuerzos por lograr un estándar común siguen siendo un desafío teniendo en cuenta que, de acuerdo al tipo de objeto que se almacene en los repositorios, se eligen unos u otros. Se destacan el estándar Dublin Core para cualquier tipo de objeto digital y el Learning Object Metadata para los objetos digitales específicamente de aprendizaje. En algunos casos resulta engorrosa la cantidad de metadatos a completar, en otros parece incompleto, por lo cual resulta necesario buscar nuevas

opciones. El presente trabajo realiza un relevamiento de las iniciativas relacionadas con la interoperabilidad, los estándares de metadatos más utilizados; y se propone una posible solución, de lo que se considera un tema transversal, la definición de metadatos.

Palabras Clave

Repositorios institucionales, interoperabilidad, metadatos.

1. Introducción

Los repositorios institucionales surgen como parte de la iniciativa de Acceso Abierto (AA), proporcionando una herramienta al área científico-tecnológica que le permite la difusión de sus trabajos a un menor costo que la manera tradicional, además de acortar los tiempos entre la finalización de un trabajo académico y su publicación en una revista. Un factor importante a tener en cuenta en este contexto, son las restricciones que imponen las editoriales que obligan al autor a ceder sus derechos sobre su trabajo científico, afectando seriamente la divulgación del mismo en forma gratuita. Asimismo, en las comunidades científicas se da mayor importancia a la publicación en revistas de alto impacto (generalmente con políticas *pay per view* o de AA línea dorada) que a aquellas que ofrecen mayor acceso a los trabajos (AA línea verde).

De esta manera surge la Iniciativa de Archivos Abiertos (OAI de su sigla en inglés Open Archives Initiative), cuyos pilares son: la consolidación de los archivos académicos con alcance mundial, acceso libre a los archivos (como mínimo a los metadatos), interfaces

similares para los archivos y proveedores de servicios y protocolo de bajo nivel/implementación sencilla (se basa en HTTP, XML, DC)

La primera iniciativa del movimiento OAI se dio en Santa Fe (Nueva México, 1999), el cual tenía como objetivo proponer una solución para la recolección de datos. Este primer paso permitiría a los proveedores de e-prints publicar sus metadatos a través de una interfaz estándar. En esta convención, los aspectos técnicos que se tuvieron en cuenta fueron: formato para los metadatos, un protocolo basado en el antiguo *Dienst*¹ y un sistema de identificación. Este movimiento, despertó interés en otras disciplinas que exigieron una revisión de los acuerdos realizados. También fue necesario trabajar sobre la credibilidad ya que esta iniciativa no contaba con una estructura organizativa que la avale, haciendo dudar a un profesional si adoptar o no sus recomendaciones. De allí, surge la necesidad de obtener respaldo. Es así que en el año 2000, la Digital Library Federation y la Coalition of Networked Information de los Estados Unidos confirman su apoyo a la misma. Uno de los principales temas en la convención de Santa Fe fue la recolección de metadatos, que permite a los proveedores de datos exponer sus metadatos para que otros repositorios puedan acceder a los objetos digitales. La versión más desarrollada del protocolo Open Archives Initiative - Protocol for Metadata Harvesting (OAI - PMH) versión 1.0 [1] fue publicada en el año 2001. Este fue un primer intento hacia la interoperabilidad entre repositorios.

La interoperabilidad es una de las principales características que hará posible la implementación del Acceso Abierto. Según el informe COAR “el valor real de los repositorios recae en el potencial de interconexión para crear una red de repositorios, que puede proveer un acceso unificado a los resultados de la investigación y son (re) utilizados por las máquinas y los investigadores. Sin embargo, para lograr este potencial, es necesario contar con interoperabilidad” [2].

De acuerdo a Rodríguez la interoperabilidad es necesaria para [3]:

- Conectar: diferentes repositorios entre sí; repositorios con otros sistemas de información tales como los sistemas de administración de e-learning (LMS de su sigla en inglés Learning Management System) y la transferencia de objetos digitales y/o metadatos.
- Crear nuevos servicios y herramientas

Las principales áreas y temas abordados por la interoperabilidad, son redes de repositorios, cosecha de metadatos, estadísticas de uso, transferencia entre

sistemas, identificación única de autor, identificación persistente y gestión de objetos compuestos tal como se muestran en la Figura 1 [4].

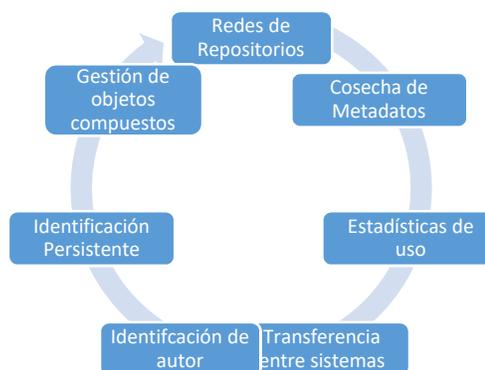


Fig. 1: Áreas de servicio relacionados con la interoperabilidad

Las *redes de repositorios* abordan la problemática de interconectar repositorios entre sí de manera de ver a la red como un gran almacén de objetos digitales. En relación a este tema se resaltan las iniciativas de Driver (Digital Repository Infrastructure Vision for European Research) que es un proyecto europeo que propone una infraestructura para repositorios digitales (esta iniciativa ha definido su propia guía como una herramienta de buenas prácticas para la implementación de un repositorio); y OpenAire (Open Access Infrastructure for Research en Europe) que también es un proyecto europeo que proporciona una orientación para poner en práctica la gestión de los datos según las demandas de acceso abierto de la Comisión Europea, éstas directrices se basan en las propuestas por Driver. A nivel latinoamericano La Referencia es la red de repositorios que reúne a los nodos de cada país socio; y a nivel nacional el Sistema Nacional de Repositorios Digitales (SNRD) propone la adhesión al mismo a través del cumplimiento de sus directrices basadas en estándares internacionales.

La *cosecha de metadatos* afronta la forma de recolectar nuevos recursos de diferentes repositorios a partir de la descripción de sus metadatos, de esta forma, los procesos de búsqueda deben coincidir en qué metadatos utilizar y cuál es el significado de los mismos. En este sentido el protocolo ampliamente utilizado y aceptado como un estándar en todos los repositorios es el OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) utilizado tanto por proveedores de datos como de servicios, los primeros para exponer sus metadatos para ser recolectados y los otros para utilizar los metadatos recolectados como servicio adicional.

En cuanto a la *gestión de objetos compuestos* el proyecto que trabaja sobre esta área es el Open Archives Initiative – Object Reuse and Exchange (OAI-ORE). Los

1

<http://www.cs.cornell.edu/cdlrg/dienst/protocols/DienstProtocol.htm>

objetos compuestos son aquellos objetos digitales multimediales (texto, imagen, video y datos). Esta iniciativa está basada en la arquitectura de la World Wide Web y construida sobre el modelo de RDF que utiliza tripletas para la descripción de sus objetos; y define estándares para la descripción y el intercambio de recursos web.

Relacionados con las áreas de servicios, y específicamente con la interoperabilidad, la última versión del informe del COAR sobre la nueva generación de repositorios plantea la definición de user stories, entre las que podemos mencionar [5]:

- Definir metadatos que permitan describir recursos educativos, como así también descubrir identificadores para recursos educativos, este último relacionado directamente con la utilización de URI HTTP persistentes
- Visualizar los derechos de uso, haciendo referencia a que debe estar explícito mediante la utilización de URI las condiciones de uso de los recursos educativos, tal como se realiza con las licencias Creative Commons.

Las user stories describen las funcionalidades prioritarias para los repositorios, teniendo en cuenta que el usuario final puede ser tanto humano como otros repositorios.

A continuación, se profundiza sobre el concepto de interoperabilidad y sus niveles; además de un análisis más detallado de algunas de estas iniciativas que se consideran de mayor relevancia; agregando además una introducción a los principales estándares de metadatos, que son necesarios para el correcto funcionamiento de la mayoría de las iniciativas antes mencionadas; concluyendo con los nuevos desafíos que propone el tema de interoperabilidad.

2. Marco Teórico

A fin de poder contextualizar los diferentes conceptos que se deben tener en cuenta al momento de analizar la interoperabilidad de repositorios institucionales, se propone las siguientes subsecciones de: interoperabilidad, estándares de metadatos, protocolo de recolección de metadatos, directrices y proyectos de redes de repositorios. Lo que permitirá unificar los criterios y conceptos sobre los cuales se hará el correspondiente análisis del funcionamiento actual de estos servicios en las redes federadas.

2.1. Interoperabilidad

Como se mencionó en la sección anterior, se puede definir la interoperabilidad de diferentes maneras. Según lo establecido en la I Cumbre Europea de

Interoperabilidad en la iAdministración [6] la interoperabilidad es “un elemento multidimensional, que integra los aspectos técnicos, semántico, organizativos, jurídico y cultural, exigiendo la existencia de equipos humanos especializados y multidisciplinares dentro de las administraciones y el fomento de grupos de trabajo interadministrativos y de órganos de composición mixta pública-privada de interoperabilidad”. Al buscar la interoperabilidad se encuentra que deben relacionarse diferentes organizaciones, que cuentan con diversidad de políticas, cultura y significado de los conceptos. Por lo cual, el trabajo se realiza entre personas especialistas en diferentes temas que aportan su know-how para posibilitar la interoperabilidad.

Como concepto general de interoperabilidad en este trabajo se adopta el siguiente: “la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada” [7]. Luego, llevando el concepto de interoperabilidad a los RI se puede mencionar a [3], donde la interoperabilidad se define como “la habilidad de los sistemas para comunicarse con otros intercambiando información, metadatos, y objetos digitales entre ellos con un ida y vuelta en un formato utilizable”.

Ahora bien, se pueden considerar diferentes niveles de interoperabilidad y la propuesta de Pérez y col. [8] distingue entre:

- Codificación: representación en caracteres.
- Léxico: representación en palabras o símbolos.
- Sintáctico: oraciones estructuradas o fórmulas o afirmaciones.
- Semántico: significado proposicional de la representación.
- Semiótico: significado pragmático de la representación o su significado en su contexto

Las tres primeras pueden agruparse como interoperabilidad sintáctica; y desde el punto de vista de la web semántica puede ser representada a través de la utilización de XML y RDF; y las dos últimas como interoperabilidad semántica.

Ahora bien, para el caso específico de los RI, según Garrido Arenas y col. [9] los niveles de interoperabilidad son:

- Infraestructura: mediante la utilización de protocolos, tales como ISO-OSI (Modelo de referencia de interconexión de sistemas abierto)² y TCP/IP (Protocolo de Control de Transmisión/Protocolo de Internet)³ para llevar a cabo el intercambio de datos.
- Sintaxis: dotar a los sistemas de información para que puedan leer datos de otros sistemas similares,

²<https://www.iso.org/ics/35.100/x/>

³ <http://www.tcpipguide.com/>

permitiendo obtener una representación que pueda ser compatible.

- Estructura: existencia de modelos lógicos comunes que permita a los sistemas de información comunicarse entre sí a través de protocolos.
- Semántica: capacidad de los sistemas de información de tener un entendimiento común de los términos que se intercambiarán.

Si se toma la clasificación propuesta por la IEEE [7] es más acotada aún, ya que se propone simplemente la *interoperabilidad sintáctica* haciendo referencia a la capacidad de dos sistemas para que puedan establecer una comunicación e intercambiar información, lo cual podrán realizarlo a través de protocolos de comunicación y transferencia, codificación de caracteres y formato de datos; y la *interoperabilidad semántica* haciendo referencia a proveer al receptor de los mecanismos necesarios para interpretar correctamente la información recibida de manera automática, sin intervención humana. En este tipo de interoperabilidad debe tenerse en cuenta los formatos de metadatos, los vocabularios controlados, ontologías y directrices de interoperabilidad.

Si se analiza las dos últimas propuestas de niveles de interoperabilidad, se puede ver que coinciden en sintáctica y semántica ya que las otras opciones están incorporadas en esa categorización más general.

Para lograr la interoperabilidad, tal cual lo menciona Garrido Arenas y col. [9] se necesitan estándares, directrices y protocolos como pilares, de allí que surgen las diferentes iniciativas para poder llevar a cabo este objetivo.

2.2. Estándares de metadatos

Si bien, dentro de las áreas de servicios que se mencionan en los párrafos anteriores, no se encuentran explícitamente el uso de los estándares de metadatos, los proyectos como OAI-PMH, SWORD, DRIVER, OpenAire, CRIS-OAR entre otros, además de los user stories planteados en la nueva generación de repositorios, necesitan de los metadatos para poder llevar a cabo sus objetivos. Por lo cual se considera la adopción del estándar de metadatos un aspecto crucial al momento de la implementación de un repositorio. Si bien, en la mayoría de los repositorios según Sandobal Verón y col. [10] el estándar de metadatos más utilizado es el Dublin Core no cualificado, la utilización de un estándar u otro depende de los objetivos que persigue la institución y el tipo de objetos digitales que están disponibles en los repositorios.

Los metadatos son descripciones sobre datos, que permiten obtener mayor información sobre el objeto digital, y además poder recuperar información

almacenada teniendo en cuenta los criterios de búsquedas seleccionados. Asimismo, Senso y col. [11] define los metadatos como “toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso, dato u objeto que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación o interoperatividad”. Es así que surgen los diferentes estándares de metadatos más relevantes, tales como Dublin Core (DC) y Learning Object Metadata (LOM). DC es el estándar de metadatos mayormente utilizado, específicamente el no cualificado [12]. Este estándar es utilizado para la descripción de cualquier objeto digital presente en la web. Los elementos de DC poseen nombres descriptivos pretendiendo tener un entendimiento a nivel semántico de los conceptos y busca la interoperabilidad global a través de la utilización de algunos vocabularios controlados. Las etiquetas utilizadas son: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, relation, coverage y rights.

Por otro lado, el estándar LOM proporciona la definición de metadatos para objetos educativos. Entendiéndose como objeto educativo a “cualquier entidad, digital o no, susceptible de ser usada en aprendizaje, educación o formación” [13]. El estándar LOM propone la división de nueve categorías de nivel superior [14]: general (título, idioma, descripción, palabras claves, entre otros), ciclo de vida (versión, estado), meta-metadatos (identificador, contribución), técnicos (tamaño, formato, requerimientos), educativos (tipo de interactividad, tipo de recurso de enseñanza, rango de edad), derechos (costo, derecho de autor y otras restricciones), relaciones (clase, recurso), anotaciones (entidad, fecha, descripción) y clasificación (propósito, descripción, palabras claves).

La incorporación de nuevos estándares de metadatos surge en la búsqueda de un equilibrio entre la generalidad con los que describe los objetos digitales el estándar DC y la enorme cantidad de etiquetas que utiliza LOM. Así comienzan a utilizarse estándares como Common European Research Information Format (CERIF), Metadata Object Description Schema (MODS), Machine-Readable Cataloging (MARC), y más recientemente DataCite(International Data Citation).

En particular CERIF⁴ es un modelo de datos relacional considerado como estándar internacional que incluye todos los aspectos de un proyecto de investigación, desde los nombres e identificadores de los proyectos, personas, organizaciones, publicaciones, entidades financiadoras, indicadores de impacto, equipamiento, entre otros. Las etiquetas utilizadas en CERIF se muestran en la Figura 2.

⁴<http://eurocris.org/cerif/feature-tour/cerif-16>

Entity	URL label
cfProject	/projects
cfPerson	/persons
cfOrgUnit	/orgunits
cfResultPublication	/publications
cfResultProduct	/products
cfResultPatent	/patents
cfFunding	/fundings
cfService	/services
cfFacility	/facilities
cfEquipment	/equipments
cfMedium	/media
cfIndicator	/indicators
cfMeasurement	/measurements
cfEvent	/events
cfPAddr	/postaladdresses
cfEAddr	/electronicaddresses
cfGeoBBox	/geoboxes
cfCitation	/citations
cfCV	/cvs
cfPrize	/prizes
cfQualification	/qualifications
cfExpertiseAndSkills	/expertiseandskills

Fig. 2: Lista de nombres válidos para entidades CERIF. Fuente: <http://dspacecris.eurocris.org>

En el caso de MODS⁵ es utilizado principalmente para la definición de elementos bibliográficos. Entre los elementos de alto nivel se puede mencionar: abstract, condición de acceso, clasificación, extensión, género, identificador, idioma, nombre, ítem relacionado, área, tabla de contenidos, audiencia destinataria, tipo de recurso, entre otros. Cada uno de estos elementos se subdivide en otros de mayor detalle, y difieren en nivel de especificidad de acuerdo al elemento a describir, así por ejemplo el elemento “idioma”, se subdivide hasta llegar a elegir el estándar en que se definirá el idioma del elemento, entre los cuales se puede seleccionar: rfc3066 (Identificación de Idiomas)⁶, iso-639-2b (Listado de Códigos de Idiomas)⁷, iso-639-3(Tercer Parte del Listado de Identificación de Idiomas, utiliza 3 letras como identificadores únicos de todos los lenguajes conocidos)⁸, rfc-4646 (Etiquetas para la identificación de idiomas, 2005)⁹ y rfc5646 (Etiquetas para la identificación de idiomas, 2009)¹⁰.

MARC¹¹ es un estándar muy conocido a nivel bibliográfico que se compone de tres elementos: estructura del registro (donde utiliza el estándar ISO 2709), designación del contenido y el contenido de los registros. Así se tiene un nivel de Información General, la

cual se divide en: nombres personales, nombres corporativos, nombre de reuniones y títulos uniformes, los cuales se estructuran según se muestra en la Figura 3.

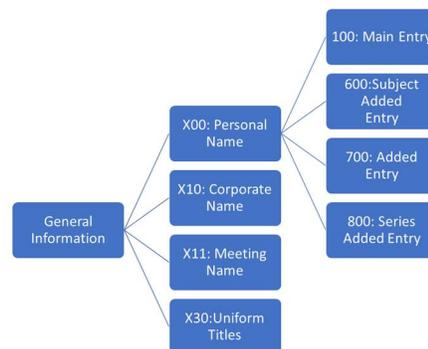


Fig. 3: Estructura de etiquetas de MARC21 para el ítem Información General

Recientemente, se ha puesto en consideración la utilización del estándar DataCite[15] que pretende proporcionar un listado de metadatos que permitan la identificación precisa y consistente de un recurso que pueda ser citado y recuperado. Entre los metadatos principales y obligatorios se encuentran: identificador, creador, título, editor, año de publicación y tipo de recurso. Estas etiquetas se subdividen dando mayor especificidad a los datos, así por ejemplo creador o autor se subdivide en (Figura 4):

- creatorName: que a la vez se subdivide en familyName (apellido) y givenName (nombre de pila).
- nameIdentifier, subdividido en nameIdentifierScheme (en este caso podría utilizarse algún esquema como ORCID) y SchemeUri (el URI del esquema utilizado en el punto anterior)

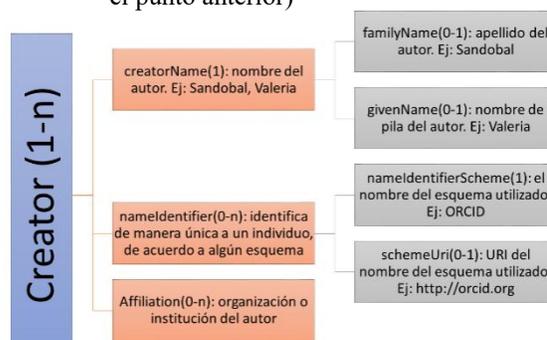


Fig. 4: Estructura e instancias de etiquetas de DataCite para el ítem de Creador

Las otras etiquetas obligatorias son: identificador (se recomienda utilizar el DOI como identificador único del recurso), título (en el caso del título se incluye además el tipo de título como ser: título alternativo, subtítulo, título

⁵<https://www.loc.gov/standards/mods/>

⁶ <https://www.rfc-editor.org/info/rfc3066>

⁷https://www.loc.gov/standards/iso639-2/php/code_list.php

⁸ <http://www-01.sil.org/iso639-3/default.asp>

⁹ <https://www.rfc-editor.org/pdf/rfc4046.txt.pdf>

¹⁰ <https://www.rfc-editor.org/info/rfc5646>

¹¹ <https://www.loc.gov/marc/>

traducido y otros), editorial, año de publicación y tipo de recurso (donde se indica una descripción del recurso y el tipo general a través de un vocabulario compartido donde los valores posibles son: audiovisual, colección, dataset, imagen, recurso interactivo, modelo, software, workflow, entre otros).

2.3. Protocolo para la recolección de Metadatos (OAI-PMH)

El protocolo OAI-PMH provee un marco de interoperabilidad basada en la recopilación de metadatos, y es la principal iniciativa en el área de cosecha de metadatos [1]. Hay dos clases de participantes en el marco de trabajo OAI-PMH:

- *Proveedores de datos*: administran sistemas que soportan el protocolo como una forma de exponer los metadatos
- *Proveedores de servicio*: utilizan los metadatos recolectados a través del protocolo como base para la creación de valor en los servicios que ofrecen.

Para realizar su operatoria el protocolo OAI-PMH utiliza peticiones utilizando los métodos get y post del protocolo http. Las peticiones posibles que un cliente puede realizar a un servidor son:

- *GetRecord*: recupera un registro concreto. Para poder responder es necesario los parámetros de identificador único del registro solicitado y el formato bibliográfico en el que se debe devolver.
- *Identify*: se lo utiliza cuando quiere obtenerse información sobre el servidor tales como: nombre, versión del protocolo, dirección del administrador, entre otros.
- *ListIdentifiers*: recupera los encabezamientos de los registros. Puede incluir el rango de fechas entre los cuales se solicita los encabezados.
- *ListRecords*: recupera los registros completos, pudiéndose incluir un rango de fecha entre los que se solicita los registros.
- *ListSets*: permite la recuperación de un conjunto de registros, los cuales son creados opcionalmente por el servidor para facilitar una recuperación selectiva de los registros. Los conjuntos pueden ser listas simples o estructuras jerárquicas.
- *ListMetadataFormats*: devuelve la lista de formatos bibliográficos que utiliza el servidor.

El protocolo OAI-PMH utiliza conceptos tales como:

- *Cosechador*: es una aplicación cliente que utiliza las peticiones OAI-PMH. Un cosechador es utilizado por un proveedor de servicio para obtener metadatos de los repositorios.

- *Repositorio*: un repositorio es un servidor que puede procesar las peticiones OAI-PMH. Un repositorio es administrado por un proveedor de datos.
- *Ítem*: es un contenedor que almacena o genera dinámicamente metadatos sobre un único recurso en múltiples formatos cada uno de los cuales puede ser cosechado por el protocolo OAI-PMH. Cada ítem cuenta con un identificador único dentro del repositorio del cual es parte.
- *Identificador único*: es un identificador que permite la individualización unívoca del ítem dentro del repositorio. El formato del identificador se corresponde con la sintaxis de URI.
- *Registro*: un registro es un metadato expresado en un único formato. El registro es una respuesta OAI-PMH a la solicitud de metadatos de un ítem. El registro se envía en XML organizada en las siguientes partes:

header	metadata	about
Unique identifier	Mínimo en formato DC no calificado.	Puede contener las declaraciones de derecho y de procedencia
timestamp		
setSpec(cero o más)		
status (optional)		

Tabla 1: Registro enviado como respuesta de una solicitud OAI-PMH

- *Set*: es una construcción opcional para agrupar elementos con el propósito de permitir la recolección selectiva

2.4. Directrices

Las directrices son un conjunto de normas que establece un marco de trabajo para que dos sistemas puedan interactuar de forma correcta.

2.4.1. Directrices Driver 2.0

Las directrices Driver pretenden llegar a la interoperabilidad en dos capas, la sintáctica (utilización de OAI-PMH y OAI-DC) y semántica (utilización de vocabularios), según lo establece [16]. Las directrices DRIVER se centran en 5 puntos

- *Colecciones*: es obligatorio utilizar “sets” que definan las colecciones de texto completo.
- *Metadatos*: se definieron algunas características obligatorias y otras recomendadas para saltar las diferentes interpretaciones que pueden surgir de la implementación de Dublin Core.
- *Implementación del protocolo OAI-PMH*: se definieron algunas características obligatorias y

otras recomendadas para solucionar las diferentes interpretaciones de la implementación del repositorio local.

- **Prácticas recomendadas:** en este punto se trata de proponer soluciones a problemas comunes a los que se pueden enfrentar los administradores de repositorios al momento de implementarlos. No son obligatorias, pero surgen de la experiencia previa. En este sentido se recomiendan las conversiones de los tipos Driver 1.0, tales como e-print, términos de versiones comunes, versiones de grupo de trabajo técnico, Journal Article Versions (JAV) a Driver 2.0; entre otras consideraciones.
- **Vocabulario y semántica:** se recomienda la utilización del espacio de nombre info:eu-repo para asignar URI a esquemas de identificadores sin URI. Esto permite tener presencia web, por lo cual tiene significado. También se menciona la identificación de autor, que se denomina DAI (Digital Author Identifier), el cual debe ser unívoco para cada autor. El formato recomendado es utilizar el International Standard for Name Identification (ISNI); combinando el DAI (dado por el RI) con su autoridad. Se indican también las recomendaciones para el vocabulario de tipo de publicaciones.

2.4.2. Directrices OpenAire

El objetivo principal de las directrices OpenAire [17] es proporcionar una orientación a los administradores de repositorios en relación al cumplimiento de las demandas de Acceso Abierto de la Comisión Europea. OpenAire proporciona 3 guías a saber: (i) Guía para administradores de repositorios de artículos científicos-académicos, (ii) Guía para administradores de repositorios de datos científicos, y (iii) Guía para administradores de CRIS (Current Research Information System)

La guía en la cual se hará referencia en este trabajo es la relacionada con los repositorios de artículos científicos académicos. En su última versión, 3.0, incluye como principales puntos la utilización de nuevos elementos para indicar identificadores alternativos, relaciones con otras publicaciones y relaciones para búsqueda de datasets; y recomendaciones de cómo utilizar los elementos de DC que han sido heredados de las directrices Driver.

OpenAire utiliza el protocolo OAI-PMH v2.0 para la recolección de los metadatos de las publicaciones. Para que los registros sean recolectados, OpenAire necesita el uso del set OAI. Este conjunto debe tener las siguientes características: *setName:OpenAire*; *setSpec:openaire*.

Las publicaciones que se insertan en el conjunto *OpenAire* deben ajustarse al menos a uno de los siguientes criterios: (i) deben estar habilitados para acceso abierto (texto completo sin restricciones de acceso), (ii) ser el resultado de un proyecto de investigación, (iii) contar con un identificador para el proyecto.

En relación al formato de metadato OpenAire utiliza OAI-DC. La sintaxis de DC toma la forma de URI y se define como un nombre de espacio info:eu-repo. En OpenAire los campos pueden ser de uso:

- **Obligatorio:** el campo debe estar siempre presente, un campo vacío no es permitido. El campo *Access Level*, en DC *rights*, identifica valores tomados de info:eu-repo-Access-Terms vocabulary; donde los posibles valores son: acceso cerrado, acceso embargado (por un período de tiempo), acceso restringido y acceso abierto.
- **Obligatorio cuando es aplicable:** el valor del campo debe ser obligatorio dependiendo del tipo de objeto digital. Como ejemplo de este tipo de campo podemos mencionar a *Embargo end date*, en DC se utiliza el identificador *date*; y se utiliza solo si el valor que toma el metadato *Access Level* es *acceso embargado* y el formato de fecha que se utiliza es YYYY-MM-DD, correspondiente a la ISO 8601.
- **Recomendado:** el uso del campo es recomendado por OpenAire lo que se considera permite una mayor descripción del objeto digital; un ejemplo de este tipo de campo podría ser *license condition*, en DC *rights*, ya que establece las condiciones en que el objeto digital puede utilizarse. Se recomienda para mayor claridad el uso de los servicios de derecho de Creative Commons [19].
- **Opcional:** los valores de este campo son útiles para brindar información complementaria sobre el objeto digital.

2.4.3. Directrices Sistema Nacional de Repositorios Digitales (SNRD)

Las directrices del sistema nacional de repositorios digitales (SNRD) surgen a partir de la promulgación de la ley N° 26899 de Creación de Repositorios Digitales en Argentina [18]. A partir de la misma, las instituciones comienzan a trabajar en la implementación de Repositorios Institucionales, lo que lleva a la necesidad de establecer reglas que definan lineamientos generales. Es así, que el SNRD tiene como objetivo principal

“conformar una red interoperable de repositorios digitales en ciencia y tecnología, a partir del establecimiento de políticas, estándares y protocolos comunes a todos los integrantes del Sistema” [19].

Los repositorios digitales que se adhieran a la aplicación de las directrices SNRD, estarán en condiciones de formar parte de redes internacionales, asegurando la interoperabilidad entre los mismos, teniendo en cuenta que aplican estándares utilizados internacionalmente.

Las directrices SNRD se basan en las “Directrices DRIVER2.0. Directrices para proveedores de contenido-Exposición de recursos textuales con el protocolo OAI-PMH”. También han sido analizadas las:

- Directrices OpenAIRE 1.1
- Directrices de la Biblioteca Digital Colombiana “Modelo de Interoperabilidad para BDCOL”
- “Directrices LUCIS/MODS” propuesta de la Biblioteca Digital de la Universidad Nacional de Cuyo.

Las directrices SNRD establecen un Set SNRD que permite obtener información en formato DC, los cuales serán cosechados mediante el protocolo OAI-PMH. El set está formado por: *setName*: Sistema Nacional de Repositorios Digitales y *setSpec*: snrd.

Los elementos que establecen las directrices SNRD están basadas en el estándar DC no cualificado y el uso de elementos mantendrá la nomenclatura de DRIVER; es decir: (i) obligatorio: el elemento siempre debe estar presente; (ii) obligatorio si es aplicable: si se puede obtener el elemento debe estar presente; (iii) recomendado: se recomienda el uso del valor de la etiqueta; y (iv) opcional: queda a criterio del repositorio la utilización o no de la etiqueta, generalmente proporcionan mayor información sobre el objeto digital.

En las directrices SNRD se establecen también la utilización de vocabularios controlados y diferentes estándares para la codificación de los metadatos. En este sentido se recomienda para:

- *dc:type*: términos DRIVER, los términos acordados por el SNRD y su correspondiente versión (la cual es recomendada, mientras que los dos primeros son obligatorios)
- *dc:format*: lista registrada de IANA de tipos de medios de Internet (tipo MIME) 6.
- *dc:creator*, *dc:contributor*: AACR2r
- *dc:language*: ISO 639-3
- *dc:date*: ISO 8601[W3CDDTF]
- *dc:coverage*: ISO 3166 para Países, DCMI BOX para áreas, Período DCMI para períodos de tiempo

2.5. Proyecto de Redes de Repositorios

En relación a los Proyectos de Redes de Repositorios pueden mencionarse las realizadas por La Referencia y OpenAire. El primero una red federada de repositorios institucionales de publicaciones científicas a nivel Latinoamericano, y el segundo un repositorio cosechador a nivel global que obtiene su contenido de repositorios institucionales, de portales nacionales y revistas electrónicas.

2.5.1. La Referencia

La Red Federada de Repositorios Institucionales de Publicaciones Científicas “La Referencia” [20] cuenta con la adhesión de países como Argentina, Brasil, Colombia, Chile, Ecuador, El Salvador, México y Perú. Entre sus objetivos se encuentra: (i) *Actualizar las recomendaciones para el control de calidad de sus metadatos en relación a las directrices Driver, presentar observaciones relacionadas con el uso del protocolo OAI-PMH; elementos a considerar de OpenAIRE y las recomendaciones específicas que las directrices especifican*, (ii) *Facilitar la adopción por los nodos nacionales e informar a los repositorios de la región de los metadatos obligatorios y recomendados, así como delinear los aspectos a profundizar en el futuro*.

La Referencia realiza la cosecha de aquellos nodos nacionales adheridos. Es así, que un repositorio institucional en particular deberá estar adherido al nodo nacional para a su vez poder ser cosechado por La Referencia.

Entre las principales recomendaciones para los nodos nacionales se encuentran:

- El campo *dc:format* queda como recomendado
- Pasan de recomendados a obligatorio si es aplicable los campos: *description*, *subject*, *language* y *publisher*.
- El campo *contributor* se incluye como recomendado, donde en los casos de director o supervisor de tesis debe ser obligatorio.
- Los campos *relation*, *coverage* y *audience* son opcionales.
- El campo *dc:right* se vuelve obligatorio; donde los artículos con los valores *restrictedAccess* y *closedAccess* no son cosechados por La Referencia

Algunas consideraciones generales que resultan de vital importancia para La Referencia tienen que ver con los campos *ProjectID*, *AuthorID*, *EmbargoEnd* e *Identificación de Fondos* están sujetos a acuerdos globales de interoperabilidad, especialmente con OpenAire.

Otra cuestión a resaltar en la cosecha de La Referencia es la identificación del repositorio e

institución de origen, que surge como una iniciativa regional que permitiría aumentar la visibilidad, generar mejores estadísticas y búsquedas. Al no estar definida en las directrices Driver ni OpenAire, es un metadato no obligatorio y que depende de la decisión de los nodos nacionales.

2.5.2. OpenAire

OpenAire[21] nace como un proyecto de implementación de una infraestructura electrónica y mecanismos de soporte para la identificación, depósito, acceso y seguimiento de artículos financiados por la FP7(Programa de Financiamiento de la Unión Europea para la investigación y la innovación 2007-2013, actualmente se denomina Horizon 2020) y el ERC (Electronic Research Collection).

Entre sus objetivos se encuentra el trabajar con comunidades temáticas para explorar los requisitos, prácticas, incentivos, flujos de trabajo, modelos de datos y tecnologías para depositar, acceder y manipular conjuntos de datos de investigación de diversas formas en combinación con publicaciones de investigación. En este sentido, realiza acciones que permitan tener la infraestructura adecuada para dar soporte al almacenamiento, la gestión y el acceso a datos científicos. Haciendo especial énfasis en la relación de las producciones científicas con los proyectos de investigación, con el conjunto de datos en que se basan las investigaciones, publicaciones relacionadas, la afiliación institucional, como así también las métricas.

Las directrices de OpenAire se basan en las directrices Driver, como se ha mencionado en secciones anteriores, con algunas adecuaciones que fueron oportunamente explicadas.

Un repositorio para poder ser cosechado debe estar registrado en OpenDOAR¹², luego pasar por un *validador de OpenAire*¹³ que sirve para verificar la compatibilidad entre el repositorio y las directrices OpenAire, y como último paso registrar el repositorio en OpenAire. De esta manera cualquier repositorio institucional que cumpla con estos requisitos podrá ser cosechado por OpenAire. La frecuencia con la que OpenAire cosecha repositorios es una vez por semana, generando informes para los administradores de los repositorios.

3. Metodología y resultados

A fin de corroborar la implementación de los diferentes aspectos que hacen a la interoperabilidad se realizó un análisis pormenorizado de las redes de

repositorios. Las que se han tomado como referencia son: a nivel nacional Sistema Nacional de Repositorios Digitales (SNRD), a nivel latinoamericano La Referencia y a nivel europeo Open Aire.

Este análisis toma como referencia informe del COAR del año 2015[22] donde menciona que resulta “cada vez más relevante para la comunidad la adopción de *metadatos comunes*, identificadores (tanto para autores, instituciones, organizaciones que financian investigaciones y publicaciones), *vocabularios* y *taxonomías*”. En este informe se han identificado problemas de interoperabilidad que aún deben solucionarse clasificándolos según su relevancia y complejidad. La clasificación propuesta para ambos casos es: alta, moderada y baja. En este sentido se ha identificado como de relevancia alta y complejidad moderada el uso de formatos de metadatos adicionales; y de relevancia y complejidad alta la utilización de metadatos de calidad.

3.1. Implementación del protocolo OAI-OMH

La implementación de este protocolo es evidente en las redes de repositorios tomadas como referencia, así como también en los repositorios de origen que son cosechados por los nodos principales. Este protocolo es necesario para obtener mayor visibilidad de las publicaciones realizadas por los autores en sus repositorios institucionales. La implementación del protocolo OAI-PMH supone la incorporación de una interoperabilidad a nivel sintáctico, ya que cada repositorio tiene su política de cosecha mencionando los requerimientos obligatorios para poder llevarla a cabo.

3.2. Adopción de directrices y estándares de metadatos

Como bien se menciona anteriormente el COAR ha definido como relevancia alta y complejidad moderada la adopción de estándares de metadatos. Algunas de las iniciativas para paliar este problema de interoperabilidad es la definición e implementación de directrices. En este sentido, podemos mencionar a nivel internacional OpenAire y Driver 2.0 son las que se utilizan en la mayoría de los repositorios, buscando homogeneizar la aplicación de directrices. Ahora bien, desde la latinoamerica la red La Referencia establece sus políticas tanto de metadatos como de cosecha, pero no define una directriz en particular, simplemente establece condiciones que deben cumplirse para que los repositorios puedan ser cosechados. Luego, cada uno de los países latinoamericanos establecen sus directrices, así por ejemplo en Perú se tiene Directrices para el

¹² <http://www.opendoar.org/>

¹³ <https://www.openaire.eu/validator/>

procesamiento de información en los repositorios institucionales. CONCYTEC¹⁴; y en particular Argentina cuenta con las Directrices establecidas por el SNRD¹⁵. Si bien, estas directrices establecen los lineamientos a seguir, la totalidad de ellas se basa en las directrices Driver 2.0. La implementación de esta directriz hace que la mayoría de los repositorios utilicen metadatos correspondientes al estándar DC, agregando en algunos casos metadatos particulares. Tal es el caso de las directrices SNRD que agrega dentro su tipo de documento un metadato que relaciona los tipos de documento establecido por Driver con los términos específicos de SNRD. Por otro lado, las directrices de Perú antes mencionada agrega como parte de sus metadatos una serie que pretende especificar de manera detallada las tesis de los diferentes niveles académicos, éstas son: *thesis.degree.level*, *thesis.degree.grantor*, *thesis.degree.name*, *thesis.degree.discipline* y *thesis.degree.program*.

3.3. Pruebas de búsquedas en las redes federadas La Referencia y OpenAire y SNRD

Teniendo en cuenta las redes de repositorios descriptas en las secciones anteriores, se realizaron búsquedas que permitan realizar un análisis de las semejanzas y diferencias, y de las dificultades al momento de la búsqueda de artículos científicos-académicos. Como así también comprobar la calidad de los metadatos utilizados en los diferentes repositorios, teniendo en cuenta que esto ha sido definido como de alta relevancia y complejidad al momento de definir problemas de interoperabilidad por el COAR.

Al realizar una *búsqueda por autor* en las 3 redes dan resultados diferentes en cuanto a cantidad (ver figuras 5, 6 y 7), es así que:

- *SNRD*: da como resultado más de 300 OD (figura 5), en algunos casos se encuentran repetidos mostrando los diferentes repositorios en los que se encuentra alojado, por ejemplo, un OD alojado en CIDISI y en CIC-Digital.



Fig. 5: Resultado de la búsqueda realizada en SNRD

- *La Referencia*: muestra una cantidad menor significativa de OD encontrados con el mismo criterio de búsqueda (figura 6), esto puede deberse a dos razones: por un lado, La Referencia sólo cosecha artículos académicos, tesis de maestrías y doctorados y reportes (o informes de investigación) dejando de lado por ejemplo los OD del tipo objeto de conferencia; y por otro lado unifica los OD indicando los enlaces de repositorios en los que se encuentra almacenados.



Fig. 6: Resultado de la búsqueda realizada en La Referencia

- *OpenAire*: los resultados mostrados son un intermedio en cuanto a cantidad de los expuestos por el SNRD y La Referencia (figura 7). Esto puede deberse en cuanto a la menor cantidad en relación al SNRD que OpenAire unifica en una sola muestra los OD alojados en diferentes repositorios indicando las fuentes de donde los ha obtenido. Para el caso del La Referencia puede deberse a que OpenAire cosecha todo tipo de OD incluido los objetos de conferencia, como así también que OpenAire cosecha desde repositorios

14

<http://portal.concytec.gob.pe/images/stories/images2013/portal/areas-institucion/dsic/directrices-procesamiento-informacion-repositorios-institucionales.pdf>

15

http://repositorios.mincyt.gob.ar/pdfs/Directrices_SNRD_2013.pdf

institucionales y La Referencia desde nodos nacionales, por lo cual podría haber repositorios que pudieran no estar adheridos al mismo; y por último el alcance de estas redes de repositorios, mientras el primero es nacional, el segundo es latinoamericano y este último es europeo, pero con cosecha a nivel internacional.

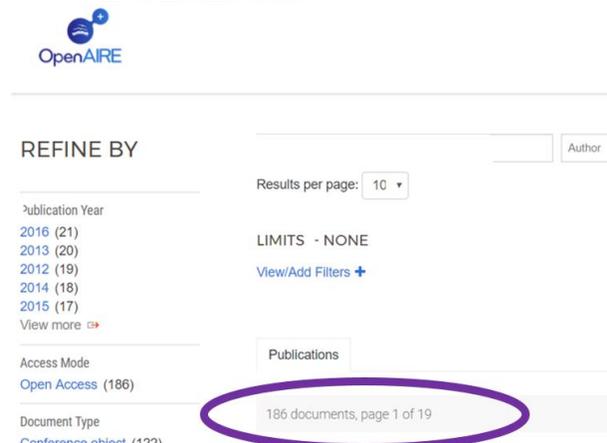


Fig. 7: Resultado de la búsqueda realizada en OpenAire

La **búsqueda por título** muestra similitudes en los resultados, por un lado, en OpenAire se muestra una sola opción con dos enlaces al mismo repositorio, pero diferente tipo de OD (artículo y objeto de conferencia). Por su lado, La Referencia muestra un resultado con tres enlaces, dos que coinciden con OpenAire mismo repositorio distinto tipo de OD; y uno más que corresponde a un journal. Por último, el resultado que expone es de tres OD, dos de ellos al mismo repositorio, pero diferente tipo de OD y uno más que pertenece a un repositorio diferente a los dos primeros.

En el caso de búsquedas específicas como por fuente, editor, materia, país entre otros, difiere en relación a la red de repositorio que se utilice. En el caso de La Referencia agrega la opción de realizar búsquedas por país teniendo en cuenta que es un cosechador a nivel latinoamericano. A diferencia de OpenAire donde no es posible especificar los tipos de búsquedas antes mencionado, dando la posibilidad de elegir entre opciones como: publicaciones, datos de investigación, personas, proyectos, organizaciones.

4. Conclusiones

En las secciones anteriores se ha realizado un relevamiento y análisis de los principales proyectos que se están llevando a cabo en el área de interoperabilidad.

Todas estas iniciativas pretenden de una forma local o global llevar soluciones relacionadas con la interoperabilidad a nivel sintáctico y la posibilidad de intercambiar información entre sistemas y entre repositorios. Ahora bien, en los proyectos mencionados se ha tratado muy someramente el aspecto semántico de la interoperabilidad, donde se pretende que cada metadato que describe un objeto digital sea entendido por todos los sistemas sin tener en cuenta el estándar de metadatos implementado.

Como se mencionó en secciones anteriores el informe del COAR, y en relación a los metadatos, mencionar que existe una necesidad de definir nuevos estándares de metadatos que resulten más convenientes para la descripción de los OA. Lo que queda reflejado en las búsquedas realizadas donde quiere realizar por una temática específica no permiten definir de manera correcta la misma, obteniendo resultados que no responden al significado de la palabra que se utilizó. Si bien DC es el estándar más utilizado y el obligatorio para la implementación OAI-PMH (o al menos exportar a este formato), este presenta cierta vaguedad en la interpretación de algunas etiquetas. Algunos estándares que se mencionan como posibles para subsanar estas falencias son: MODS, METS, MARC, CERIF y DataCite entre otros.

La adopción de nuevos estándares de metadatos, si bien tienen el objetivo de describir con mayor especificidad el OA, lleva a una diversidad de estándares utilizados y con ellos la dificultad de interoperabilidad a nivel semántico de los mismos. En este sentido, se considera que la definición a nivel ontológico de metadatos para la descripción de OA, teniendo en cuenta las diferentes fuentes de datos heterogéneas que están presentes en los repositorios, ayudaría a solucionar el problema proporcionando el significado común a los mismos. Donde se tome como base el estándar DC, teniendo en cuenta su utilización masiva, y agregando nuevos metadatos que describan de manera significativa a los OA, pero que a la vez puedan ser entendidos por otros repositorios. Aún más considerando que la próxima generación de repositorios tiene entre sus objetivos prestar servicios que puedan ser utilizados tanto por usuarios humanos como máquina.

Referencias

- [1] The Open Archive Initiative Protocol for Metadata Harvesting. Protocol Version 2.0. Disponible en: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#Introduction>
- [2] El caso de Interoperabilidad para Repositorios de Acceso Abierto. Julio 2011, v1. Disponible en: <https://www.coar-repositories.org/es/activities/repository->

[interoperability/coar-interoperability-project/a-case-for-interoperability/](#)

[3] Rodrigues, E. Challenges and Opportunities of Interoperability for Open Access Repositories. China Open Access Week. Beijing, Octubre 2012.

[4] The Current State of Open Access Repository Interoperability. COAR. Octubre, 2012. V2. Disponible en:

<https://www.coar-repositories.org/es/activities/repository-interoperability/coar-interoperability-project/the-current-state-of-open-access-repository-interoperability-2012/>

[5] Next Generation Repositories, COAR. Febrero, 2017. Disponible en:

<https://www.coar-repositories.org/files/COAR-Next-Generation-Repositories-February-7-2017.pdf>

[6] I Cumbre Europea de Interoperabilidad en la Administración. Valencia, Noviembre 2006. Disponible en:

<http://www.educoas.org/RestrictedSites/Curso1/Newsletter-Septiembre07/paratenerencuenta28.html>. Consulta: 05/08/2015

[7] Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York: 1990

[8] Euzenat, J. Towards a principled approach to semantic interoperability. En: A. Gómez-Pérez; M. Grüniger; H. Stuckenschmidt y M. Uschold (Eds.), Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence. Seattle, US, August 4–5. 2001

[9] Garrido Arenas, H.; Lisowska Navarro, M. ¿Estamos preparados para trabajar en red? Interoperabilidad: desafíos para la región latinoamericana. XX ISTE (Ibero-American Science & Technology Education Consortium) General Assembly. Puebla, México. Marzo 2014

[10] Sandobal Verón, 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. II Jornadas de Investigación en Ingeniería del NEA y Países Limítrofes. Junio 2012. ISBN: 978-950-42-0142-7

[11] Senso, José A., and Antonio de la Rosa Piñero. "El concepto de metadato. Algo más que descripción de recursos electrónicos." *Ciência da Informação* 32.2 (2003): 95-106.

[12] Dublin Core Metadata Element Set, Version 1.0: Reference Description (1998). Disponible en: <http://dublincore.org/documents/1998/09/dces/#>. Consultado: 10 de Agosto de 2015.

[13] Estándar para Metadatos de Objetos Educativos. Institute of Electrical and Electronics Engineers, Inc. Park Avenue New York, NY 10016-5997 USA. 2002

[14] IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata.

Version 1.3 Final Specification. Disponible en: http://www.imsglobal.org/metadata/mdv1p3/imsmd_best_v1p3.html. 2006.

[15] DataCite Metadata Working Group. (2016). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. DataCite e.V. <http://doi.org/10.5438/0012>.

[16] Directrices Driver 2.0. Directrices de proveedores de contenido. Noviembre, 2008

[17] OpenAire Guidelines for Literature Repositories v3.0. Disponible en:

<https://guidelines.openaire.eu/en/latest/literature/index.html>. Abril, 2013

[18] Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos. Noviembre, 2013. Texto definitivo disponible en:

<http://repositorios.mincyt.gob.ar/recursos.php>.

[19] Directrices SNRD. Directrices para proveedores de contenido del Sistema Nacional de Repositorios Digitales. Ministerio de Ciencia, Tecnología e Innovación Productiva. Mayo 2013

[20] La Referencia. Red Federada de Repositorios Institucionales de Publicaciones Científicas. Metadatos y Políticas de cosecha. Mayo de 2015.

[21] OpenAire. Open Access Infrastructure for research in Europe. Framework Programme 7(2007-2013)

[22] COAR, Working Group 2, Repository Interoperability. COAR Roadmap Future Direction for Repository Interoperability. Febrero, 2015. Disponible en:

https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf