

Los elementos en la representación de recursos en los repositorios institucionales: una revisión bibliográfica

Jose Texier

Universidad Nacional Experimental del Tachira (UNET), Venezuela
Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata
(SEDICI), Argentina
CONICET, Argentina
jtexier@unet.edu.ve

Marisa R. De Giusti

Servicio de Difusión de la Creación Intelectual, Universidad Nacional de La Plata
(SEDICI), Argentina
marisa.degiusti@sedici.unlp.edu.ar

Resumen

Esta revisión se centra en identificar cómo la literatura estudia los problemas existentes en la Representación de Recursos (RR) de los Repositorios Institucionales (RI). La RR es el proceso de registrar en forma persistente un conjunto de datos (metadatos) como síntesis y reemplazo del objeto "real" para poder identificarlo, recuperarlo y distribuirlo. La RR se define por ciertos elementos: recursos, esquemas de metadatos, almacenamiento y catalogación. Por su parte, los RI se basan en procesos funcionales de acuerdo con el material depositado y la norma ISO 14.721: carga (*ingest*), almacenamiento (*storage*), catalogación (*cataloging*), indización (*indexing*), búsqueda (*search engine*) y navegación (*browsing*). Los resultados de esta revisión muestran que identificar los problemas presentes en estos elementos y procesos funcionales no es objeto de estudio por parte de los investigadores, lo cual desemboca en un área de vacancia en este dominio, para resolver parte de los problemas presentes en los RI a partir de los elementos y procesos funcionales.

Palabras clave: Repositorios Institucionales, Representación de Recursos, ISO 14.721, Problemas

Abstract

This review focuses on identifying how the literature studies the existing problems in the Resource Representation (RR) of Institutional Repositories (IR). RR is a process of recording in a persistent manner a set of data (metadata) as a synthesis and replacement of the "real" object, to allow its identification, retrieval and dissemination. RR is defined by certain elements: resources, metadata schema, storage and cataloging. On the other hand, IRs are based on functional processes according to the material that is deposited and the ISO 14.721 standard: *ingest*, *storage*, *cataloging*, *indexing*, *search engine* and *browsing*. The results of this review show that identifying the problems found in these elements and functional processes is not a subject of study for the researchers, which leads to a vacant area in this field.

Keywords: Institutional Repositories, Resource Representation, ISO 14.721, problems.

1. Introducción

En la última década los repositorios institucionales (RI) han crecido [1] y representan una fuente de información digital especializada, organizada y accesible para los usuarios de diversas áreas. Los RI son sistemas informáticos dedicados a gestionar los trabajos científicos y académicos de diversas instituciones de forma libre y gratuita [2], además, están relacionados con los ideales y objetivos del Acceso Abierto [3], [4], y contribuyen a repensar los procesos de publicación de artículos científicos [5]. De la misma forma, las Bibliotecas Digitales comenzaron su evolución a partir de 1990 [2] y se fueron consolidando en el mundo científico con el pasar de los años, hasta entrelazarse con el concepto y funcionalidad de los RI, por ello, en el contexto de este trabajo un Repositorio Institucional es una Biblioteca Digital y una Biblioteca Digital es un Repositorio Institucional, gracias a que ambos ofrecen servicios similares y el uso de cada término depende del ámbito donde se aplica y, por ende, de los recursos con los que se desean trabajar [2].

Los diferentes trabajos relevados sobre el funcionamiento de un RI [6]–[11] y según las recomendaciones de la norma ISO 14.721 [10] (también conocida como el modelo OAIS), todo repositorio debe tener los siguientes procesos funcionales de acuerdo con el material depositado: carga (*ingest*), almacenamiento (*storage*), catalogación (*cataloging*), indización (*indexing*), búsqueda (*search engine*) y navegación (*browsing*). A partir del funcionamiento del depósito los procesos funcionales son: preservación y administración. Por ello, la Representación de Recursos (RR) dentro de un RI está definida por los procesos funcionales relacionados con el material almacenado y con el proceso de registrar en forma persistente de un conjunto de datos como síntesis y reemplazo del objeto "real" para poder identificarlo, recuperarlo y distribuirlo por parte de los usuarios [7]. Cuando se hace referencia a recursos, se refiere a objetos físicos o digitales que se describen a partir de la enumeración de un conjunto de datos específicos (llamados metadatos) que se distinguen entre otros objetos [7].

El concepto de metadatos no es algo nuevo, antes de la aparición de Internet se usaba en la catalogación de libros y revistas para normalizar la información de manera que se pudiera recuperar organizadamente. En el ámbito de las Ciencias de la Información, los metadatos se emplean para referirse a registros de recursos de información disponibles [13]. En otras palabras, los metadatos son datos que caracterizan otros datos, es decir, es información estructurada que describe, explica y/o localiza un recurso de información para poder identificarlo, recuperarlo, utilizarlo, administrarlo o preservarlo de una manera más clara y sistemática. Para la representación de metadatos se han desarrollado distintos modelos, esquemas, formatos o estándares, que si bien comparten una sintaxis y estructura de la información en XML, difieren atendiendo a los propósitos de la información que describen [14].

A partir de lo explicado anteriormente se definen cuatro elementos en una representación de recursos dentro de los repositorios institucionales [7], [12], ejes de esta revisión bibliográfica:

- tipología de los recursos,
- esquemas de metadatos,
- almacenamiento, y,
- catalogación representada por los vocabularios controlados, tesauros y las entidades abstractas (elementos con información descriptiva propia) tales como autores, instituciones y revistas.

Por tanto, el objetivo de esta revisión bibliográfica es conocer cómo la literatura científica analiza el problema de la representación de recursos en los repositorios institucionales, visto como un todo, en otras palabras, encontrar soluciones que involucren los seis procesos funcionales de un RI –de acuerdo con el material depositado– y los cuatro elementos de una RR. Para ello, la revisión se ha organizado

de la siguiente manera: en la segunda sección se expone la metodología usada para llevar a cabo la revisión; en la tercera sección se describen los resultados; en la cuarta sección se expone un análisis, la discusión de los resultados y el aporte del trabajo presentado; y, finalmente, en la quinta sección se presentan unas conclusiones.

2. Metodología

La revisión bibliográfica sistemática consiste en identificar, evaluar e interpretar todas las investigaciones relevantes posibles de manera rigurosa para responder una pregunta, área particular de investigación o fenómeno de interés [15]. Para el desarrollo de esta revisión se tomaron como base algunas pautas de la literatura médica [16], [17] y los criterios definidos por Kitchenham en el 2004 [15]:

- **Pregunta de investigación:** La pregunta que guió la revisión bibliográfica fue determinar cómo la literatura aborda el tema de la representación de recursos (elementos) dentro de los repositorios institucionales (procesos funcionales) considerando las recomendaciones de la Norma ISO 14.721.
- **Evaluación de la estrategia de búsqueda:** Tal evaluación se organizó, según las directrices generales PICOC [18], que consideran la eficacia desde cinco puntos de vista:
 - Población: la representación de recursos en el dominio LIS.
 - Intervención: los elementos de una representación de recursos y los procesos funcionales en los repositorios institucionales.
 - Comparación: de acuerdo con las recomendaciones de la norma ISO 14.721, se analizan y comparan los problemas relacionados con los elementos y los procesos funcionales.
 - Resultado: en las búsquedas no se limita el tipo de resultados según criterios establecidos, ya que eran necesaria toda la información disponible en el dominio de estudio.
 - Contexto: no se aplicó ninguna restricción.
- **Estrategia de búsqueda y criterios:** Se definieron dos grupos de búsquedas en la base de datos bibliográfica Scopus, por ser la más grande base de datos actualmente (REFERENCIAR JISC-ADAT), de acuerdo con los lineamientos de la norma ISO 14.721. El primer grupo de búsquedas, llamada “problemas en los elementos” (PE), se orientó sobre los problemas existentes en cada uno de los cuatro elementos (incluidos los subelementos) de una representación de recursos: recursos, metadatos, almacenamiento y catalogación, siempre en el contexto del dominio LIS (Library & Information Science). El segundo bloque de búsquedas, conocido como “problemas en los procesos” (PP), se basó en los problemas existentes en los seis procesos funcionales de los repositorios y su relación con los elementos de la representación de recursos. Los criterios de búsqueda en Scopus para ambos grupos son comunes (PE y PP), a saber: tipos de documentos a tomar en cuenta, *article* y *review*; campos de selección en la base de datos son: título, resumen y palabras claves; y, áreas temáticas de los artículos, “*Computer Science*” y “*Social Science*”. No hay restricción referente a los años de publicación y todas las búsquedas se realizaron el 5 de noviembre del 2013. En el grupo PE las 10 búsquedas están basadas en cada uno de los elementos y subelementos de la representación de recursos, los resultados se restringieron de acuerdo con la aparición de la cadena “*digital library*” como palabra clave en las publicaciones. Los términos de descriptores truncados usados fueron:
 1. (types* AND resource*) AND problem*
 2. metadata AND problem*
 3. storage AND problem*
 4. catalog* AND problem*
 5. “controlled vocabular*” AND problem*

6. thesaurus AND problem*
7. "abstract entiti*" AND problem*
8. author AND problem*
9. institution AND problem*
10. journal AND problem*

El segundo grupo (PP) consistió en 24 búsquedas para identificar los problemas existentes entre los seis procesos funcionales definidos y los cuatro elementos de una representación de recursos en el dominio LIS. Los siguientes seis términos de descriptores truncados fueron la base de las búsquedas, y cada uno se relacionó con los cuatro elementos de la RR:

1. ingest*
2. storage
3. catalog*
4. indexing
5. "search engine"
6. browsing

- **Extracción y síntesis de datos:** Los resultados obtenidos en las 34 búsquedas (10 del grupos PE y 24 de grupo PP), se exportaron desde Scopus como archivos CSV (*comma-separated values*), que luego fueron incorporados a Google Refine [19], herramienta que ofrece funcionalidades adicionales a los gestores de hojas de cálculo como LibreOffice Calc o Excel. Estos archivos de resultados y el proceso detallado de las búsquedas realizadas se encuentran en un proyecto GitHub [20].

3. Resultados

3.1. Grupo Problemas en los Elementos (PE)

En la Tabla 1 se muestra los artículos encontrados que corresponden a los cuatro elementos y seis sub-elementos de la representación de recursos. En la tercera columna ("Art. con términos DL & IR") se exponen los resultados con la presencia de los términos *Digital Library* o *Institutional Repository*. En la última columna ("Art. limitados por áreas CS & SocS") se representan los artículos encontrados cuando la búsqueda se limita a las áreas *Computer Science* y *Social Science*. Esta última columna representa un primer resultado según los criterios planteados (248 artículos), para seleccionar el corpus que será analizado en la cuarta sección.

N	Elementos – Términos que exponen problemas	Art. con términos DL & IR	Art. limitados por áreas CS & SocS
1	types OR resources	507	130
2	metadata	221	37
3	storage	181	45
4	cataloging	69	28
5	"controlled vocabular*"	11	2
6	thesaurus	20	9
7	"abstract entities"	0	0
8	author	164	40
9	institution	100	32
10	journal	77	26
	TOTAL - unión	979	248

Tabla 1. Problemas en los elementos de una RR

3.2. Grupo Problemas en los Procesos (PP)

En la Tabla 2 se muestra la cantidad de artículos encontrados que identifican los problemas presentes en los RI de acuerdo con los seis procesos funcionales de un repositorio en relación con los cuatro elementos de una representación de recursos. La última columna (“*intersección*”) indica si hay algún artículo que trate el problema en los cuatro elementos de una representación de recursos para un proceso funcional específico de un RI.

Procesos funcionales	Resources	Metadata	Storage	Cataloging	Intersección
<i>ingest</i>	1	3	3	3	0
<i>storage</i>	44	31	-	29	2
<i>cataloging</i>	7	6	3	6	0
<i>indexing</i>	20	23	23	24	0
<i>“search engine”</i>	35	22	4	24	0
<i>browsing</i>	17	9	8	6	0

Tabla 2. Relación de los procesos y los elementos

Los dos artículos obtenidos al hacer la intersección de los elementos con el proceso *storage* son:

- Integrating chemistry scholarship with web architectures, grid computing and semantic web [21].
- Data for the future The German project "Co-operative development of a long-term digital information archive" [22].

El primero de ellos fue presentado en una conferencia, razón por la cual y para mantener el criterio de búsquedas del grupo PE (solamente artículos en revistas y revisiones bibliográficas), fue descartado. En cambio, el segundo artículo ya está incluido dentro del grupo PE.

3.3. Consideraciones generales

Se pueden mencionar algunos resultados generales tomando en cuenta los 248 artículos obtenidos en las búsquedas con respecto a: los autores, los años de publicación y las revistas.

- **Los autores:** En la Tabla 3 se observan 5 autores –de los 558 encontrados en los 248 artículos– que cuentan con el mayor número de publicaciones, es decir, entre 6 y 3 artículos cada uno. Inferimos que este dato refleja la existencia de unas líneas de investigación como el modelo conceptual 5S por Goncalves, Fox y Laender [23], las bibliotecas digitales en la educación universitaria por Fox [24], [25], deduplicación de autores por Goncalves, Ferreira y Laender [26], y sistemas de recuperación por Herrera-Viedma [27].

Autores	Cant. Art.
Goncalves Marcos A.	6
Laender Alberto H.F	4
Ferreira Anderson A.	3
Fox Edward A.	3
Herrera-Viedma Enrique	3

Tabla 3. Autores con mayor presencia en los resultados

- **Años de publicación:** La Figura 1 muestra una línea del tiempo en qué momentos y cuántas publicaciones se reportan con respecto al tema. Destacan los años 2007 y 2012 con 26 y 25 artículos, esto evidencia lo novedosa que es esta área de investigación.

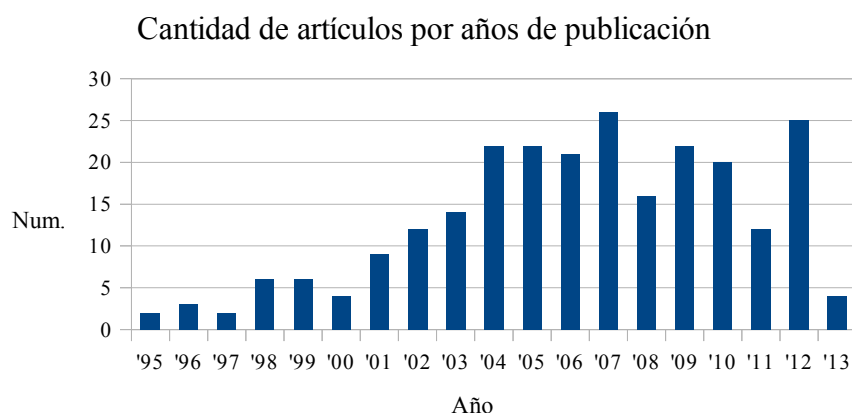


Figura 1. Años de publicación de los resultados

- **Revistas:** La Tabla 4 muestra las primeros 13 revistas de un total de 110, ordenadas en forma descendiente con la cantidad de artículos publicados a partir de los 248 artículos encontrados.

Revistas	Cant. Art.
Electronic Library	44
Program	17
OCLC Systems and Services	15
Library Hi Tech	14
Online Information Review	14
Journal of the American Society for Information Science and Technology	13
Science and Technology Libraries	11
Computer Physics Communications	9
International Journal on Digital Libraries	8
Aslib Proceedings: New Information Perspectives	7
IEEE Transactions on Knowledge and Data Engineering	6
Journal of Digital Information	6
Journal of Library Metadata	6

Tabla 4. Revistas con mayor presencia en los resultados

3.4. Corpus

El propósito de este trabajo es analizar la literatura encontrada que expone los diferentes problemas existentes en la representación de recursos dentro de un repositorio institucional. Para ello, se hallaron 248 publicaciones que abordan al menos un elemento de la RR. No obstante, al hacer una intersección entre los elementos y subelementos, a través del proyecto de Google Refine, se encontró un artículo en el que están presentes los cuatro elementos principales de la representación (ver Tabla 5).

Grupos	Cant. Art.
Un elemento	194
Dos elementos	45
Tres elementos	8

Cuatro elementos	1
TOTAL	248

Tabla 5. Relación de elementos

Nuestro interés reside en conocer la representación de recursos como un todo, por tanto, en la Tabla 5 se evidencia un área de vacancia en el dominio LIS porque solamente se muestra una publicación con ese fin [22]. De manera tal que en la Tabla 6 se muestra el corpus definitivo (9 artículos) que tomamos para nuestro estudio a partir de la selección de las publicaciones enfocadas en tres y cuatro elementos.

N	Artículo o Review	Elemento o subelementos
1	Data for the future The German project "Co-operative development of a long-term digital information archive" [22]	4 elementos: resources, metadata, storage, cataloging
2	Towards accessibility to digital cultural materials: An FRBRized approach [28]	3 elementos: resources, metadata, cataloging
3	The growth of electronic journals in libraries: Access and management issues and solutions [29]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>cataloging</u>
4	BibPro: A citation parser based on sequence alignment [30]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>cataloging</u>
5	Digital library development: Identifying sources of content for developing countries with special reference to India [31]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>cataloging</u>
6	From digital library to institutional repository: A brief look at one library's path [32]	3 elementos: <u>metadata</u> , <u>storage</u> , <u>cataloging</u>
7	Help features in digital libraries: Types, formats, presentation styles, and problems [33]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>cataloging</u>
8	Provision of digital preservation metadata: A role for ONIX? [34]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>storage</u>
9	Subject Access: Conceptual Models, Functional Requirements, and Empirical Data [35]	3 elementos: <u>resources</u> , <u>metadata</u> , <u>cataloging</u>

Tabla 6. Corpus obtenido

4. Análisis y Discusión

El corpus objeto de análisis fue de 9 artículos a partir de los 248 artículos encontrados. Este primer hallazgo evidencia un área poco estudiada en el dominio LIS, puesto que tanto en artículos de investigación como revisiones bibliográficas, no se aborda el tema de la representación de recursos dentro de los repositorios institucionales como un todo donde se involucren los seis procesos funcionales de un RI y los cuatros elementos de una RR.

De los artículos obtenidos, se descartaron tres por las razones que se explican en cada caso. Un primer trabajo fue el de Xie [33], porque no trata ningún problema relacionado con el propósito de esta revisión. Los trabajos de Chen *et al.* [30] y de Buelhler *et al.* [32], porque se enfocan en la integración de un repositorio institucional de citaciones de artículos y de bibliotecas digitales, respectivamente, provenientes de diversas plataformas de software consolidadas. Estos artículos resaltan entre sus problemas y soluciones, algunos relacionados con los elementos de una RR y procesos funcionales de los RI, pero no cumplen con el criterio de estudiar al menos 3 de los elementos de una RR.

Los 6 artículos que formaron parte del corpus definitivo fueron aquellos que exponen problemas relacionados con el propósito definido para este estudio, por ejemplo: diversidad de soluciones tecnológicas, tratamiento de diferentes tipologías de recursos, esquemas de metadatos,

recomendaciones de almacenamientos de recursos, la preservación de recursos, la recomendaciones del modelo OAIS y la aplicación de modelos conceptuales para solucionar como un todo el problema de la RR tales como el modelo FRBR (*Functional Requirements for Bibliographic Records*), FRAD (*Functional Requirements for Authority Data*) y FRSAD (*Functional Requirements for Subject Authority Data*) [22], [28], [29], [31], [34], [35].

Los ejes de análisis, de acuerdo con los elementos de RR y procesos funcionales de los RI, que agruparon los planteamientos de las publicaciones halladas se centran en:

1. *Recursos*: el trabajo de Altenhöner [22] menciona que los objetos digitales deben verse y tratarse como unos bitstreams, según la norma ISO, por el contrario, en los trabajos de Weng et al. [28] y Mischo et al. [29] estudian la diversidad de recursos culturales y electrónicos, adaptados a esquemas de metadatos conocidos como MARC o METS. De igual manera, Jeevant [31] explica el tratamiento de recursos académicos y científicos digitalizados y Brindley et al. [34] se centran en la migración del recurso libro. En cambio, Zavalina estudia el recursos en forma general y adaptado a la familia de modelos FRBR [35]. En resumen, se identifican un gran número de recursos que deben ser parte de un RI, tales como: artículos, revisiones, proceedings, monografías, tesis, datasets, documentos administrativos, documentos de gobierno, reportes técnicos, etc. De esta manera un RI tiene que adaptarse a las tipologías existentes y a los nuevos tipos de recursos que deben salir con el pasar del tiempo.
2. *Esquemas de metadatos*: en el corpus se evidencian varios esquemas tradicionales como soluciones a los RI. Entre los más mencionados está el esquema METS estudiados por Altenhöner [22] y Brindley et al. [34]; y el MARC por Weng et al. [28] y Mischo et al. [29]. Adicionalmente, los autores Altenhöner [22], Mischo et al. [29], Jeevant [31] y Brindley et al. [34] usan esquemas de desarrollo propio y de propósito general, por ejemplo Dublin Core. Mención especial merece el trabajo Brindley et al. [31] porque expone el tema de la preservación digital y recomiendan el uso del esquema PREMIS. Todos estos esquemas deben ser permitidos por los RI para evitar la pérdida de información y hacerlos interoperables con otros RI. También los repositorios deben ser capaces de adaptarse a futuros esquemas de metadatos que se establezcan por recomendaciones o por ser estándares de facto.
3. *Almacenamiento*: en el trabajo de Jeevant [31] se recomienda cómo debe ser proceso para gestionar la persistencia de los metadatos y el objeto digital, y Brindley et al. [34] trabaja directamente con el paradigma de base de datos relacional. Cabe destacar que los trabajos de Altenhöner [22] y Brindley et al. [34] mencionan las recomendaciones de la norma ISO 14.721. Por ello, estos trabajos se enfocan en solucionar la persistencia de la información a través del modelo y en garantizar la recuperación de la información de forma normal (a través de consultas) y ante desastres no previstos (a través de copias espejo o distribuidas). Asimismo, recomiendan el uso de identificadores de recursos persistentes y la ejecución de tareas de revisión y modificación de los recursos para mejorar la integridad y calidad de la información, ya que los depósitos provienen de diferentes fuentes y vías, y del uso correcto de controles bibliográficos.
4. *Catalogación*: los trabajos encontrados recomiendan diferentes vías para garantizar la normalización de la información ingresada y almacenada dentro de un RI. Weng et al. [28] se centran en la recomendación del modelo FRBR para catalogar los diferentes tipos de recursos, Mischo et al. [29] sugieren el uso estricto de controles bibliográficos como el análisis documental formal y de contenido, y el trabajo de Jeevant [31] recomienda el uso de una guía para catalogar basada en los principios básicos de la disciplina, Asimismo, Zavalina [35] recomienda el uso de los modelos de la familia FRBR (FRBR, FRAD y FRSAD) y la aplicación del código de catalogación RDA (*Resource Description and Access*) para no usar las AACR2 (segunda edición de las reglas de catalogación *Anglo-American-Cataloguing*) o sus

predecesores, y Zavalina enfoca el problema de la catalogación a través del acceso por materia a partir de dos actores: los usuarios que buscan la información en los sistemas de repositorios y los profesionales de la información que analizan y crean los metadatos de los recursos. Por tanto, estos trabajos se encargan de facilitar los puntos de acceso para que los usuarios localicen y recuperen la información indizada, que por lo general, se centra en los campos título, autor y materia, y en algunos casos, permite la búsqueda a texto completo.

5. *Incorporación de recursos*: basado en el proceso funcional de ingreso de ítems a un repositorio, el trabajo de Altenhöner [22] señala que los depósitos provienen de diferentes fuentes y se realizan por diversas vías, pero no se enfoca en una solución con base en alguna norma, como sucede en el trabajo de Brindley et al. [34] quienes recomiendan mantener los criterios de la norma ISO y el uso de los paquetes de información, que para este caso serían los SIP. Situación similar ocurre en el trabajo de Jeevant [31] que solamente explican la importancia del proceso de incorporación pero no lo menciona como un problema ni una ventaja, simplemente como uno de los pasos a seguir para hacer disponible los recursos para una comunidad.
6. *Gestión de los datos*: los trabajos encontrados en cómo poblar, mantener y acceder a la información refieren en el caso de Weng et al. [28] en la aplicación del modelo FRBR y los trabajos de Altenhöner [22] y Mischo et al. [29] en usar guías de catalogación para el control de la información. De nuevo, Zavalina [35] recomienda que el manejo de los datos se centraliza sobre la familia de modelos FRBR. En síntesis, estos trabajos se enfocan en poder crear los puntos de acceso y de normalizar la información (metadatos) de los objetos digitales ante el crecimiento de los repositorios, procesos de migración o el simple uso cotidiano del mismo.
7. *Acceso*: para el usuario el acceso es importante ya que representa la puerta de entrada al repositorio. Los trabajos encontrados estudian la problemática del acceso a partir de los procesos funcionales *Indexing*, *Search* y *Browsing*. En los trabajos de Weng et al. [28] y de Zavalina [35] se estudia una solución al acceso a través de la aplicación del modelo FRBR, la cual ayuda a elaborar una buena catalogación, permitiendo que la indización de los recursos sea correcta y garantice resultados correctos para las búsquedas y navegación por los diferentes puntos de acceso definidos. Sin embargo, Zavalina incorpora la importancia de la gestión de las materias para un correcto acceso e incorpora los modelos FRAD y FRSAD [35]. Los trabajos de Mischo et al. [29] y Jeevant [31] dan pautas generales en el acceso web, pero no plantean soluciones concretas al problema.

Los siete ejes de análisis reportados surgen de los cuatro elementos de una RR y los cuatro módulos funcionales mencionados explícitamente en el modelo ISO de acuerdo con el material depositado en el RI. Los autores exponen sus problemas en estas áreas y reportan una diversidad de soluciones para cada análisis que generan dudas con respecto a si existen soluciones que puedan abarcar lo mejor de cada una de las propuestas de estos autores.

5. Conclusiones

En la segunda sección se expuso la metodología aplicada para esta revisión y las diferentes búsquedas realizadas en Scopus y el uso de Google Refine, que pueden ser encontrados en un proyecto GitHub [20], de modo que cualquier investigador pudiera replicar esta revisión. Por tanto, esta revisión permitió concluir lo siguiente:

- En los grupos de búsquedas realizadas, grupos PE y PP, se evidencia un área de vacancia en el dominio LIS, específicamente en la representación de recursos dentro de un repositorio institucional de acuerdo con los diferentes trabajos relevados y la norma ISO 14.721. Los pocos artículos con relaciones de elementos (Tabla 5) y el corpus obtenido (Tabla 6), ayudan a profundizar los diferentes problemas que se presentan en los RI como un todo y comenzar a

construir soluciones en esa misma dirección.

- La Tabla 1, refleja como la representación de recursos existe en forma amplia, pero el tema se reduce de forma considerable, un 25,33% (de 979 a 248 artículos), si el problema de RR se puntualiza en el mundo de las Ciencias de la Computación, Ciencias de la Información y Ciencias Documentales. Por ello, se infiere que la representación de recursos no es un tema exclusivo en los repositorios institucionales o bibliotecas digitales.
- La Figura 1, pone de manifiesto lo novedoso y dinámico que es el tema abordado en esta revisión. Aunque se inician las publicaciones en 1995, hay una presencia creciente y constante desde el 2001.
- La Tabla 4 muestra las revistas con más artículos encontrados, donde esas 13 revistas de un 110, poseen 170 artículos de los 248 encontrados, representando el 68,55%. Se considera un porcentaje muy alto que podría llamar la atención para hacer un estudio sobre esos resultados y el área de la representación de recursos en los repositorios institucionales.
- Finalmente, se recomienda realizar un estudio que permita relacionar el problema de la representación de recursos con los diferentes modelos conceptuales de bibliotecas digitales y repositorios institucionales que han sido mencionados a lo largo de esta revisión, a saber: modelo formal propuesto por Goncalves et al. [23], modelo de referencia OAIS [12], modelo conceptual FRBR [36], entre otros. Esto permite ver en detalle los repositorios a través de un modelo general, que pueda extraer lo mejor de cada uno de ellos sobre la base de los elementos y procesos funcionales estudiados en esta revisión.

6. Referencias

- [1] OpenDOAR, "OpenDOAR - Home Page - Directory of Open Access Repositories," 2013. [Online]. Available: <http://www.opendoar.org/>. [Accessed: 21-Oct-2013].
- [2] J. Texier, "Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior," presented at the 11th Latin American and Caribbean Conference for Engineering and Technology - 2013, Cancun, Mexico, 2013, p. 9.
- [3] B.-C. Björk and D. Solomon, "Open access versus subscription journals: a comparison of scientific impact," *BMC Medicine*, vol. 10, no. 1, p. 73, Jul. 2012.
- [4] P. Suber, "Ensuring open access for publicly funded research," *BMJ*, vol. 345, 2012.
- [5] H. A. Piwowar, R. S. Day, and D. B. Fridsma, "Sharing Detailed Research Data Is Associated with Increased Citation Rate," *PLoS ONE*, vol. 2, no. 3, p. e308, Mar. 2007.
- [6] M. Lesk, *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann, 1997.
- [7] J. Texier, M. R. De Giusti, N. Oviedo, A. J. Lira, and G. L. Villarreal, "La representación de recursos en los repositorios institucionales. El caso de estudio: SEDICI," presented at the III Conferencia de Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL) y VIII Simposio Internacional de Bibliotecas Digitales (SIBD) (Costa Rica, 2013), 2013.
- [8] C. A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL*, Feb-2003. [Online]. Available: <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>. [Accessed: 28-Oct-2013].
- [9] H. Van de Sompel, S. Payette, J. Erickson, C. Lagoze, and S. Warner, "Rethinking Scholarly Communication," *D-Lib Magazine*, vol. 10, no. 9, Sep. 2004.
- [10] E. A. Fox, M. A. Gonçalves, and R. Shen, *Theoretical Foundations for Digital Libraries. The 5S Approach*. North Carolina: Morgan & Claypool Publishers.
- [11] M. A. Gonçalves, B. L. Moreira, E. A. Fox, and L. T. Watson, "What is a good digital library? – A quality model for digital libraries," *Information Processing & Management*, vol. 43, no. 5, pp. 1416–1437, Sep. 2007.
- [12] CCSDS, *Reference Model for an Open Archival Information System (OAIS):ISO 14721*. 2012.
- [13] R. Heery, "Review of metadata formats," *Program: electronic library and information systems*, vol. 30, no. 4, pp. 345–373, Dec. 1996.
- [14] E. Méndez, "Tratamiento de los objetos de información en los archivos : retos y estándares para la descripción basada en metadatos," 2003. [Online]. Available: <http://eprints.rclis.org/handle/10760/12691#.UAAZFUEzfgM>. [Accessed: 13-Oct-2013].
- [15] B. Kitchenham, *Procedures for Performing Systematic Reviews*. 2004.

- [16] NHMRC, "How to review the evidence: systematic identification and review of the scientific literature | National Health and Medical Research Council," 1999. [Online]. Available: <http://www.nhmrc.gov.au/guidelines/publications/cp65>. [Accessed: 06-Nov-2013].
- [17] CRD, "Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews," *Undertaking Systematic Reviews of Research on Effectiveness: CRD's guidance for those carrying out or commissioning reviews*, Mar-2001. [Online]. Available: <http://www.york.ac.uk/inst/crd/crdreports.htm>. [Accessed: 06-Nov-2013].
- [18] C. Pons, R. Giandini, and G. Arevalo, "A systematic review of applying modern software engineering techniques to developing robotic systems," vol. 32, no. 1, p. 8, Apr. 2012.
- [19] Google, "Google Refine." [Online]. Available: <https://code.google.com/p/google-refine/>. [Accessed: 19-Nov-2013].
- [20] J. Texier, "ReviewRR - GitHub," 2013. [Online]. Available: <https://github.com/dantexier/ReviewRR/>. [Accessed: 06-Nov-2013].
- [21] S. K. Challa, M. Pierce, and S. Marru, "Integrating chemistry scholarship with web architectures, grid computing and semantic web," in *Gateway Computing Environments Workshop (GCE), 2010*, 2010, pp. 1–8.
- [22] R. Altenhöner, "Data for the Future: the German Project 'Co-operative Development of a Long-term Digital Information Archive' (kopal)," *Library Hi Tech*, 2006. [Online]. Available: <http://eprints.rclis.org/9199/>. [Accessed: 11-Nov-2013].
- [23] M. A. Gonçalves, E. A. Fox, L. T. Watson, and N. A. Kipp, "Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries," *ACM Trans. Inf. Syst.*, vol. 22, no. 2, pp. 270–312, Apr. 2004.
- [24] J. Pomerantz, B. M. Wildemuth, S. Yang, and E. A. Fox, "Curriculum Development for Digital Libraries," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, New York, NY, USA, 2006, pp. 175–184.
- [25] E. A. Fox, "Digital Libraries - Wikiversity." [Online]. Available: http://en.wikiversity.org/wiki/Digital_Libraries. [Accessed: 17-Nov-2013].
- [26] R. G. Cota, A. A. Ferreira, C. Nascimento, M. A. Gonçalves, and A. H. F. Laender, "An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 9, pp. 1853–1870, 2010.
- [27] C. Porcel and E. Herrera-Viedma, "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries," *Knowledge-Based Systems*, vol. 23, no. 1, pp. 32–39, Feb. 2010.
- [28] C. Weng and J. Mi, "Towards accessibility to digital cultural materials: a FRBRized approach," *OCLC Systems & Services*, vol. 22, no. 3, pp. 217–232, Jul. 2006.
- [29] W. H. Mischo, M. A. Norman, W. A. Shelburne, and M. C. Schlembach, "The Growth of Electronic Journals in Libraries," *Science & Technology Libraries*, vol. 26, no. 3–4, pp. 29–59, 2007.
- [30] C.-C. Chen, K.-H. Yang, C.-L. Chen, and J.-M. Ho, "BibPro: A Citation Parser Based on Sequence Alignment," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 2, pp. 236–250, 2012.
- [31] V. K. J. Jeevan, "Digital library development: identifying sources of content for developing countries with special reference to India," *The International Information & Library Review*, vol. 36, no. 3, pp. 185–197, Sep. 2004.
- [32] M. A. Buehler and M. S. Trauernicht, "From digital library to institutional repository: a brief look at one library's path," *OCLC Systems & Services*, vol. 23, no. 4, pp. 382–394, Nov. 2007.
- [33] H. (Iris) Xie, "Help features in digital libraries: types, formats, presentation styles, and problems," *Online Information Review*, vol. 31, no. 6, pp. 861–880, Nov. 2007.
- [34] G. Brindley, A. Muir, and S. Proberts, "Provision of digital preservation metadata: a role for ONIX?," *Program: electronic library and information systems*, vol. 38, no. 4, pp. 240–250, Dec. 2004.
- [35] O. L. Zavalina, "Subject Access: Conceptual Models, Functional Requirements, and Empirical Data," *Journal of Library Metadata*, vol. 12, no. 2–3, pp. 140–163, 2012.
- [36] FRBR, "Functional Requirements for Bibliographic Records - IFLA." 2009.