

La incorporación del registro de obras y autores de una institución: estudio de caso SEDICI-UNLP-Scopus.

Texier, José, De Giusti, Marisa Raquel, Lira, Ariel Jorge y Villarreal, Gonzalo Luján.

Cita:

Texier, José, De Giusti, Marisa Raquel, Lira, Ariel Jorge y Villarreal, Gonzalo Luján (Octubre, 2014). *La incorporación del registro de obras y autores de una institución: estudio de caso SEDICI-UNLP-Scopus. Conferencia Internacional Acceso Abierto, Preservación Digital, Interoperabilidad, Visibilidad y Datos Científicos - BIREDIAL 2014, Porto Alegre.*

Dirección estable: <https://www.aacademica.org/marisa.de.giusti/15>

ARK: <https://n2t.net/ark:/13683/ptyc/Rgp>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Conferência Internacional Acesso Aberto, preservação digital, interoperabilidade, visibilidade e dados científicos - Brasil, 2014

Eje temático: Interoperabilidad

La incorporación del registro de obras y autores de una institución: estudio de caso SEDICI-UNLP-Scopus

Texier, Jose

Universidad Nacional Experimental del Táchira (UNET), Venezuela.

Servicio de Difusión de la Creación Intelectual (SEDICI), Universidad Nacional de La Plata, Argentina.
jtexier@unet.edu.ve; dantexier@sedici.unlp.edu.ar

De Giusti, Marisa R.

Servicio de Difusión de la Creación Intelectual (SEDICI), Universidad Nacional de La Plata, Argentina y Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICPBA).

Lira, Ariel Jorge

Servicio de Difusión de la Creación Intelectual (SEDICI), Universidad Nacional de La Plata, Argentina

Villarreal, Gonzalo Luján

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

Resumen:

El Servicio de Difusión de la Creación Intelectual (SEDICI), es el repositorio institucional de la Universidad Nacional de La Plata (UNLP), cuenta con recursos de diferentes tipologías como artículos de revistas científicas, publicaciones en congresos, tesis (doctorado, maestría, especialización y pregrado), normativas y ordenanzas, libros y libros electrónicos, audios, fotografías, entre otros. El objetivo principal de este trabajo fue recuperar y dar cuenta de todas las obras producidas por autores de la UNLP, para dar difusión y visibilidad dentro de su repositorio institucional e incrementar las producciones presentes en el mismo. En julio del 2013 se inició una recopilación de todos los artículos, objetos de conferencia y artículos de revisión que se encuentran en la base de datos bibliográfica Scopus comprendidos entre 1927 y 2013. Para lograr tal objetivo, se definió un mecanismo semiautomático de incorporación de recursos desde Scopus al DSpace implementado en SEDICI. Este mecanismo fue desarrollado en sucesivas fases, que pueden resumirse en: recuperación desde Scopus, filtrado, limpieza, transformación y normalización de los metadatos, importación de autores y de los recursos, revisión, filtrado, incorporación de los PDF (si la licencia lo permite) y finalmente publicación en el repositorio. La consulta sobre Scopus finalizó el 1 de marzo de 2014 en la UNLP y fue posible gracias al servicio de pago que Elsevier brinda al Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina. Esto implica que este tipo de consultas puede ser realizada por otras instituciones públicas de investigación de la Argentina que estén bajo el paraguas del MinCyT. En resumen, el trabajo tuvo dos logros tangibles: se incorporaron a SEDICI 19.940 documentos con filiación de la UNLP y 89.903 autores de cuales al eliminar los no repetidos se redujo a 24.161 autores. Esta información permitió sumar recursos a la producción local alojada en el repositorio, mejorar la calidad de los metadatos ya almacenados e incrementar la visibilidad del repositorio y de la universidad. Por otro lado, a partir de ISSN/ESSN de los listados de revistas de DOAJ, PubMed, SherpaRomeo, SciELO y Redalyc, se relacionaron con los artículos importados desde Scopus, lo que permitió detectar que al menos un 74% de los documentos detectados admite su incorporación a texto completo en el repositorio, ya sea porque la revista es de acceso abierto, es una revista que permite preprint o postprint, o porque el autor ha pagado el arancel correspondiente para que su artículo esté en abierto.

Palabras clave: interoperabilidad, repositorios institucionales, SEDICI, DSpace, Scopus.

1. Introducción

El Servicio de Difusión de la Creación Intelectual, SEDICI [1], es el repositorio institucional de la Universidad Nacional de La Plata (UNLP) que se encuentra implementado en DSpace [2]. Para agosto del 2014, cuenta con aproximadamente 32.000 recursos de diferentes tipologías como artículos de revistas científicas, publicaciones en congresos, tesis (doctorado, maestría, especialización y pregrado), normativas y ordenanzas, libros y libros electrónicos, audios, fotografías, entre otros [3]. Los repositorios institucionales son estructuras web interoperables encargadas de recopilar, catalogar, gestionar, acceder, difundir y preservar toda la producción académica y científica de las instituciones descriptos por medio de un conjunto de datos específicos (metadatos) [4]–[7].

A mediados del año 2013, se comenzó un estudio sobre la incorporación de los recursos (artículos, objetos de conferencia y artículos de revisión) desde la base de datos bibliográfica Scopus [8] al DSpace implementado en SEDICI, a partir de todas las filiaciones de la UNLP registradas en Scopus y comprendidos en el lapso 1927-2013. El objetivo del estudio fue recuperar y dar cuenta de todas las obras producidas por autores de la UNLP en Scopus (tiene aproximadamente de 22.000 revistas [9]), para dar difusión y visibilidad dentro de su repositorio institucional e incrementar las producciones presentes en el mismo.

Para lograr tal objetivo, se definió un mecanismo semiautomático de incorporación de recursos desde Scopus a SEDICI-DSpace que puede servir de modelo a otras universidades e instituciones. Este mecanismo fue desarrollado en sucesivas fases, que pueden resumirse en: recuperación desde Scopus, filtrado, limpieza, transformación y normalización de los metadatos, importación de autores y de los recursos, revisión, filtrado, incorporación de los PDF (si la licencia lo permite) y finalmente publicación en el repositorio. La consulta sobre Scopus finalizó el 1 de marzo de 2014 en la UNLP y fue posible gracias al servicio de pago que Elsevier (editorial que gestiona Scopus) brinda al Ministerio de Ciencia, Tecnología e Innovación Productiva de la República Argentina. Esto implica que este tipo de consultas puede ser realizada por otras instituciones públicas de investigación de la Argentina que estén bajo el paraguas de dicho Ministerio.

2. Metodología

Las fases nombradas anteriormente se pueden resumir en 5 etapas:

2.1. Etapa 1 – Recuperación de metadatos

Consistió en consultar en Scopus los artículos, objetos de conferencia y artículos de revisión presentes en todas las dependencias relacionadas con la UNLP, y luego exportar los resultados obtenidos a archivos en formato CSV. La consulta y exportación finalizó en marzo del presente año, garantizando la recolección de la producción científica hasta el año 2013. Las 14 filiaciones y los artículos recopilados para cada una se observan en la Tabla 1.

Filiación	Art.
Observatorio Astronómico de La Plata	45
Centro de Química Inorgánica (CEQUINOR)	54
Instituto de Investigaciones Bioquímicas de La Plata (INIBIOLP)	98
Instituto de Botánica Carlos Spegazzini	134
Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB)	140
Instituto de Limnología Dr. Raul A. Ringuelet	139
Centro de Investigación y Desarrollo en Criotecnología de Alimentos (CIDCA)	168
Centro de Investigación y Desarrollo En Ciencias Aplicadas Dr. Jorge Juan Ronco (CINDECA)	196
Centro de Investigaciones Geológicas - UNLP	194
Centro de Investigaciones Ópticas - La Plata	438
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA	864
UNLP Facultad de Ciencias Medicas	1,023
Museo de La Plata	2,420
UNLP	14,027
TOTAL	19,940

Tabla 1. Filiaciones y artículos en Scopus

Scopus, permite exportar una serie de metadatos que en su mayoría corresponden a parte del modelo de metadatos presente en SEDICI. El modelo de metadatos que se puede exportar de Scopus es:

- *Citation information*
 - Author(s) / Formato Apellidos coma Iniciales
 - Document title
 - Year
 - Source Title
 - Volume, Issue, Pages
 - Citation count
 - Source and Document Type
- *Bibliographical information*
 - Affiliations
 - Serial identifiers (e.g. ISSN)
 - DOI
 - PubMed ID
 - Publisher
 - Editor(s)
 - Language of Original Document
 - Correspondence Address
 - Abbreviated Source Title
- *Abstract and Keywords*
 - Abstract
 - Author Keywords
 - Index Keywords
- *Fund Details*
 - Number
 - Acronym
 - Sponsor

- *References*
 - References
- *Other information*
 - Tradenames and Manufacturers
 - Accession numbers and Chemicals
 - Conference information

2.2. Etapa 2 – Normalización

Se tomaron los siguientes metadatos de los artículos exportados por Scopus y se incorporaron a Google Refine [10], lo que permitió realizar análisis, transformaciones y normalizaciones de manera flexible y rápida en dicha herramienta.

- Author(s)
- Document title
- Year
- Source Title
- Volume
- Issue
- Pages start
- Pages end
- Pages count
- Cited by Citation count
- Link Scopus
- Affiliations
- Author with Affiliations
- Abstract
- Author Keywords
- Index Keywords
- References
- Correspondence Address
- Editors
- Sponsors
- Publisher
- Conference name
- Conference date
- Conference location
- Conference code
- ISSN
- ISBN
- CODEN
- DOI
- PubMed ID
- Language of Original Document
- Abbreviated Source Title
- Document Type

- Source

El total de registros incorporados a Google Refine desde los archivos CSV obtenidos desde Scopus fueron 19.940 artículos comprendidos entre 1920 y 2013. Luego, Google Refine, permite exportar la información en archivos con formatos XML de acuerdo con plantillas que el usuario puede definir. La plantilla que se usó para exportar los 19.940 artículos fue:

```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<dublin_core schema="dc">
<doc>
<field name="cd">{{(cells["Conference date"].value)}}</field>
<field name="y1">{{jsonize(cells["Year"].value)}}</field>
<field name="ab"><![CDATA[{{(cells["Abstract"].value)}}]]></field>
<field name="p1">{{jsonize(cells["Page count"].value)}}</field>
<field name="p2">{{jsonize(cells["Page start"].value)}}-{{jsonize(cells["Page end"].value)}}</field>
<field name="la">{{(cells["Language of Original Document"].value)}}</field>
<field name="t1"><![CDATA[{{(cells["Title"].value)}}]]></field>
<field name="pu"><![CDATA[{{(cells["Publisher"].value)}}]]></field>
<field name="ty">{{(cells["Document Type"].value)}}</field>
<field name="cl">{{(cells["Conference location"].value)}}</field>
<field name="li"><![CDATA[{{(cells["Link"].value)}}]]></field>
<field name="ur"><![CDATA[{{(cells["DOI"].value)}}]]></field>
<field name="in">{{(cells["ISSN"].value)}}</field>
<field name="ib">{{(cells["ISBN"].value)}}</field>
<field name="cn"><![CDATA[{{(cells["Conference name"].value)}}]]></field>
<field name="st"><![CDATA[{{(cells["Source title"].value)}}]]></field>
<field name="vo">{{(cells["Volume"].value)}}</field>
<field name="is">{{(cells["Issue"].value)}}</field>
<field name="pm">{{(cells["PubMed ID"].value)}}</field>
<field name="k1"><![CDATA[{{join(cells["Index Keywords"].value.split("; "), ",")}}]]></field>
<field name="k1"><![CDATA[{{(cells["Index Keywords"].value.split("; "), ",")}}]]></field>
<field name="ka"><![CDATA[{{join(cells["Author Keywords"].value.split("; "), ",")}}]]></field>
<field name="ka"><![CDATA[{{(cells["Author Keywords"].value.split("; "), ",")}}]]></field>
<field name="a1">{{join(cells["Authors"].value.split(", "), ",")}}</field>
<field name="a1">{{(cells["Authors"].value.split(", "), ",")}}</field>
<field name="ed">{{join(cells["Editors"].value.split(", "), ",")}}</field>
<field name="ed">{{(cells["Editors"].value.split(", "), ",")}}</field>
<field name="su">{{(cells["SubType"].value)}}</field>
</doc>
</dublin_core>
```

Luego de exportar esta plantilla en Google Refine, se obtuvo un archivo en formato XML que contiene los 19.940 registros, el cual sirvió como entrada a un programa en Java para transformarlo en otros archivos XML con las estructuras de los esquemas de metadatos DC, SEDICI y MODS que soporta el modelo de metadatos de SEDICI-DSpace.

Se destaca, que todos estos registros se relacionaron con los listados de revistas que entregan DOAJ [11], PubMed [12], SherpaRomeo [13], SciELO [14] y Redalyc [15]. Esta relación fue posible a través de los ISSN y ESN que tiene cada uno de los documentos importados de Scopus y de los listados de dichas revistas. Las categorías obtenidas como nuevos metadatos para cada documento de Scopus, con énfasis en el Acceso Abierto (Open Access) [16], [17], son:

- **CC:** artículos que están con licencia Creative Commons, gracias a la consulta a PubMed
- **DOAJ:** artículos que pertenecen a revistas en Acceso Abierto
- **OA:** artículos en Acceso Abierto en forma parcial o total, gracias a DOAJ
- **SciELO:** artículos que la revista a la que pertenece esta en SciELO

- **Redalyc**: artículos que la revista a la que pertenece esta en Redalyc
- **Verde**¹: artículos que la revista a la que pertenece esta en SherpaRomeo con el estatus de que se puede archivar el pre-print y el post-print o versión de editor/PDF
- **Amarillo**¹: artículos que la revista a la que pertenece esta en SherpaRomeo con el estatus que se puede archivar el pre-print (ie la versión previa a la revisión por pares)
- **Blanco**¹: artículos que la revista a la que pertenece esta en SherpaRomeo con el estatus donde el archivo no está formalmente admitido
- **Azul**¹: artículos que la revista a la que pertenece esta en SherpaRomeo con el estatus donde se puede archivar el post-print (ie la versión final posterior a la revisión por pares) o versión de editor/PDF
- **Otros**: artículos sin agrupamiento

Con esta clasificación, se agruparon los 19.940 registros que se encontraban en Google Refine de la siguiente manera (Tabla 2):

Clasificación por grupos	Ítems
CC	532
Content DOAJ	319
OA	738
Scielo	787
Redalyc	84
Sherpa - Verde	9,977
Sherpa - Amarillo	1,575
Sherpa - Blanco	974
Sherpa - Azul	201
Otros	4,753
TOTAL	19,940

Tabla 2. Número de ítems por metadato

2.3. Etapa 3 – Importación de autores

La base de datos de autores de SEDICI se encuentra en una base de datos secundaria en MySQL, la cual es accedida desde el módulo de *authorities* de DSpace cada vez que se debe seleccionar un autor. Cuando se estaba haciendo la prueba de importación, se tuvo que modificar el SEDICI-Dspace para que realizara una búsqueda en la base de datos secundaria.

Para cada autor de Scopus, se mantiene sus apellidos, iniciales de su nombre y una afiliación sin normalizar porque aparece en Scopus exactamente como el autor lo colocó en el documento que aprobó la revista. La importación de estos autores se realizó antes que los documentos en DSpace, garantizando la existencia de tales autores, los cuales están clasificadas de la siguiente manera:

- Existen 84.903 autores en la base de datos recopilada de Scopus, que pueden estar repetidos.
- Ejecución de una operación para detectar duplicados en los 84.903, la lista de autores se redujo a 23.161.
- Existen 48.867 autores de los 84.903 exportados, que tienen en sus

1 Extraído de SherpaRomeo, gracias a un programa propio en Python [18]

afiliaciones “UNLP” o “La Plata”, pero se aplicó una operación similar a la anterior para detectar duplicados, ese número se redujo a 7.538.

- Los 7.538 autores, se relacionaron con los existentes en la base de datos secundaria de autores SEDICI, y se obtuvieron 3.123 autores que no coinciden, es decir, son nuevos.
- Existen 26.554 (de los 84.903) autores que no tienen en sus afiliaciones “UNLP” o “La Plata”.

Debido a la diversidad de información de los autores, principalmente enfocado en su filiación, para cada artículo incorporado y aprobado por los catalogadores de SEDICI, los autores fueron revisados manualmente e incorporados a la base de datos secundaria de SEDICI, para luego ser detectados desde SEDICI-DSpace.

2.4. Etapa 4 – Importación a DSpace de registros

DSpace provee varios mecanismos de importación de recursos que varían en complejidad y completitud de la información que representan. El usado en este trabajo es el Simple Archive Format, el cual es un formato básico que consiste en que para cada documento se genera una carpeta (items) con un archivo textual de índice, los bitstreams y con los metadatos expresados en archivos XML separados en archivos por cada formato de metadatos (DC, MODS & SEDICI) de acuerdo con el modelo de metadatos de SEDICI-DSpace. Un ejemplo de la operación a nivel de consola es:

```
./dspace import -a -e dantexier@hotmail.com -w -c 10915/25294 -s items/ -m mapfile.in
```

Se realizó la importación de prueba en el servidor de pruebas y de desarrollo sin problemas de 19.940 registros en 2241 seg. Después se actualizó el *discovery* para que DSpace reconozca los nuevos registros ingresados por la importación. Esta importación se realizó en una computadora portátil Sony VAIO con las siguientes características:

- Procesador: Intel Core i5-3337U CPU @ 1.80 GHz x 4
- Memoria RAM: 8 GB
- Disco duro de estado sólido (SSD) de 160 GB.
- Sistema operativo: Ubuntu 12.04 (precise) 64 bits y un kernel linux 3.5.0-23.generic

Otra opción para importar items en DSpace es Biblio-Transformation-Engine (BTE), pero las pruebas que se realizaron y por disponibilidad de recursos humano, se decidió continuar el proceso con la opción anterior².

Antes de realizar la importación, se tomó en cuenta el proceso de catalogación final realizado por los bibliotecarios de SEDICI, porque son ellos quienes realizan la revisión de cada uno de los metadatos importados desde Scopus, los normalizan, descargan los artículos desde los sitios web de las revistas (siempre y cuando la licencia lo permita) y asignan el licenciamiento de acuerdo con las políticas de cada una de las revistas de donde proviene el ítem. Por tanto, se agruparon los artículos en colecciones. Las colecciones, con base en las relaciones

2 Info en la Wiki de DSpace: <https://wiki.duraspace.org/pages/viewpage.action?pageId=32481931>

establecidas con las revistas comentadas anteriormente, que se crearon son:

- CC
- DOAJ
- OA
- Scielo
- SherpaRomeo
- Redalyc
- Otros

2.5. Etapa 5 – Deduplicación

Luego de la importación de autores y de los documentos, se hace una revisión y filtrado de cada uno, gracias a una tarea de curation para detectar duplicados, después si la licencia lo permite se incorporan los PDF y finalmente se autoriza la publicación en el repositorio por parte de los catalogadores. La tarea de curation se implemento usando la interfaz `AbstractCurationTask` del paquete `org.dspace.curate`. En esta interfaz se establece una conexión con SOLR a través de las clases `CommonsHttpSolrServer`, `ModifiableSolrParams`, `QueryResponse` y `SolrDocumentList`. Pero antes, se debe configurar tres `SearchHandler` para hacer búsquedas por los campos `dc.title`, `dc.description.abstract` y `sedici.creator.person`. De esta manera, se filtra la búsqueda para detectar el duplicado por la intercepción obtenidos por cada `SearchHandler` y configurando el parámetro “mm”. A continuación parte del código del archivo de configuración:

```
<requestHandler name="titulos" class="solr.SearchHandler">
  <lst name="defaults">
    <str name="echoParams">explicit</str>
    <int name="rows">10</int>
    <str name="defType">edismax</str>
    <str name="q.alt">*:*</str>
    <str name="q.op">AND</str>
    <float name="tie">0.01</float>
    <str name="qf">
      dc.title^0.9
    </str>
    <str name="mm">1<75%</str>
  </lst>
</requestHandler>
```

Luego de obtener, vía SOLR, las tres consultas de acuerdo con el título, resumen y autores, se realiza una intercepción entre ellos cuyo resultado son ítems duplicados.

3. Resultados

Los documentos incorporados en SEDICI, permitieron iniciar los procesos de eliminación de duplicados en todos los documentos del repositorio, mejoraron la calidad de los metadatos y se encontraron PDF que algunos registros existentes en SEDICI no los tenía. Además, esta información permitió sumar recursos a la producción local alojada e incrementar la visibilidad del repositorio y de la universidad.

En resumen, el trabajo tuvo dos logros tangibles: se incorporaron a SEDICI

19.940 documentos con filiación de la UNLP y 89.903 autores de cuales al eliminar los no repetidos se redujo a 24.161 autores de los cuales 3.123 son nuevos autores en SEDICI. Por otro lado, este estudio permitió detectar que al menos un 74% (Figura 1) de los documentos detectados de acuerdo con la licencia de la revista a la cual pertenece, admite su incorporación a texto completo en el repositorio, ya sea porque la revista es de Acceso Abierto, es una revista que permite preprint o postprint, o porque el autor ha pagado el arancel correspondiente para que su artículo esté en abierto. En la Tabla 3, se exponen los artículos importados desde Scopus que son candidatos a estar en Acceso Abierto y clasificados por año.

Posiblemente en OA

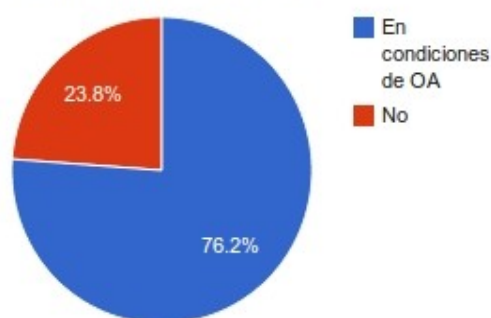


Figura 1. Artículos importados candidatos a estar en Acceso Abierto

No OA y OA por año		
Año	OA	No OA
1,920	21	1
1,927	0	1
1,929	2	0
1,930	0	1
1,932	3	0
1,934	1	0
1,939	1	0
1,944	1	1
1,945	1	0
1,946	1	0
1,947	0	1
1,949	1	0
1,952	1	0
1,953	1	0
1,954	0	1
1,955	1	1
1,956	2	2
1,958	0	1
1,959	1	0
1,960	2	0
1,961	2	1
1,962	12	3
1,963	5	8
1,964	11	3
1,965	13	8
1,966	12	6
1,967	23	3
1,968	21	7
1,969	23	6
1,970	40	7
1,971	57	12
1,972	43	12

1,973	55	30
1,974	52	26
1,975	38	29
1,976	39	26
1,977	36	33
1,978	34	24
1,979	44	22
1,980	57	23
1,981	59	39
1,982	75	43
1,983	76	47
1,984	84	62
1,985	109	47
1,986	105	46
1,987	136	69
1,988	145	69
1,989	142	64
1,990	164	68
1,991	160	70
1,992	157	57
1,993	173	75
1,994	205	108
1,995	217	81
1,996	413	119
1,997	433	145
1,998	460	145
1,999	465	153
2,000	512	140
2,001	552	179
2,002	546	168
2,003	594	149
2,004	592	173
2,005	641	200
2,006	662	225
2,007	745	214
2,008	809	264
2,009	934	234
2,010	994	246
2,011	964	240
2,012	1,013	262
2,013	1,194	253
Total	15,187	4,753

Tabla 3. Número de artículos incorporados

4. Consideraciones finales

El trabajo realizado, cumplió con el objetivo principal propuesto de recuperar y dar cuenta de todas las obras producidas por autores de la UNLP, además ayudo a incrementar la visibilidad del repositorio institucional SEDICI. La implementación de esta propuesta fue extensa y con una diversidad de opciones tecnológicas, la cual podrá mejorarse, por esta razón todo el código utilizado esta compartido en un proyecto GitHub [18].

De igual manera, se recomienda realizar la tarea de curation para detectar duplicados de forma periódica, ya que el deseo es tener la mejor calidad posible de metadatos del repositorio SEDICI.

Finalmente, se recomienda las siguientes líneas de trabajo en pro de

consolidar el objeto principal de esta investigación, ofrecer e incrementar las diferentes obras de la UNLP, y mejorar la visibilidad del repositorio:

- Simplificar el proceso de incorporación de Scopus a cualquier repositorio a través de un framework que permita la representación de dichos recursos de forma independiente, para luego adaptarlos a SEDICI o a cualquier otro repositorio.
- Replicar este estudio en otros repositorios y/o otras plataformas de software.
- Mejorar la tarea de curation deduplicación y compararla con otras alternativas de detección de duplicados.

5. Referencias

- [1] M. De Giusti, N. Oviedo, A. Lira, A. Sobrado, J. Martinez, and A. Pinto, "SEDICI – Desafíos y experiencias en la vida de un repositorio digital," *RENATA*, vol. 1, no. 2, pp. 16–33, Aug. 2011.
- [2] DSpace, "DSpace," 2014. [Online]. Available: <http://www.dspace.org/>. [Accessed: 12-Mar-2014].
- [3] M. R. De Giusti, A. Sobrado, A. J. Lira, M. M. Vila, and G. L. Villarreal, "SEDICI (Servicio de Difusión de la Creación Intelectual)," *D-Lib Magazine*, vol. 14, Sep. 2008.
- [4] J. Texier, "Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior," at the 11th Latin American and Caribbean Conference for Engineering and Technology - 2013, Cancun, Mexico, 2013.
- [5] J. Tramullas, "Bibliotecas digitales," presented at the VI Seminario de Centros de Documentación Ambiental y Espacios Naturales Protegidos, 2007.
- [6] C. L. Liew, "Digital library research 1997-2007: Organisational and people issues," *Journal of Documentation*, vol. 65, no. 2, pp. 245–266, Jun. 2009.
- [7] I. H. Witten, D. Bainbridge, and D. M. Nichols, *How to Build a Digital Library*. Morgan Kaufmann, 2003.
- [8] Elsevier, "Scopus - Welcome to Scopus," 2014. [Online]. Available: <http://www.scopus.com/>. [Accessed: 14-Aug-2014].
- [9] Elsevier, "Scopus | Elsevier," 2014. [Online]. Available: <http://www.elsevier.com/online-tools/scopus>. [Accessed: 15-Aug-2014].
- [10] OpenRefine, "OpenRefine," 2014. [Online]. Available: <http://openrefine.org/>. [Accessed: 28-May-2014].
- [11] DOAJ, "Directory of Open Access Journals," 2014. [Online]. Available: <http://doaj.org/>. [Accessed: 15-Aug-2014].
- [12] NCBI, "Home - PubMed - NCBI." [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed>. [Accessed: 15-Aug-2014].
- [13] JISC, "SHERPA/RoMEO," 2014. [Online]. Available: <http://www.sherpa.ac.uk/romeo/?la=es>. [Accessed: 15-Aug-2014].
- [14] SciELO, "SciELO - Scientific Electronic Library Online," 2014. [Online]. Available: <http://www.scielo.org/php/index.php?lang=es>. [Accessed: 15-Aug-2014].
- [15] Redalyc, "Redalyc. Latin America, the Caribbean, Spain and Portugal Scientific Journals Network.," 2014. [Online]. Available: <http://www.redalyc.org/home.oa>. [Accessed: 15-Aug-2014].
- [16] 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.
- [17] P. Suber, "Timeline of the Open Access Movement," 2009. [Online]. Available: <http://www.earlham.edu/~peters/fos/timeline.htm>. [Accessed: 31-Jan-2014].
- [18] J. Texier, "ReviewRR - GitHub," 2014. [Online]. Available: <https://github.com/dantexier/ReviewRR/>. [Accessed: 06-Nov-2014].