

Conferencia Internacional  
**BIREDIAL-ISTEC**  
17-18-19 OCTUBRE 2016

---

# Memorias

---

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ  
SISTEMA DE BIBLIOTECAS  
Biblioteca Virtual Universitaria  
San Luis Potosí, S.L.P.; México



UNIVERSIDAD DE  
COSTA RICA



UNIVERSIDAD  
NACIONAL  
DE LA PLATA



# MEMORIA DEL EVENTO

- VI CONFERENCIA INTERNACIONAL SOBRE BIBLIOTECAS Y REPOSITORIOS DIGITALES DE AMÉRICA LATINA (BIREDIAL)
- XI SIMPOSIO INTERNACIONAL DE BIBLIOTECAS DIGITALES (SIBD)

Bajo siguientes ejes temáticos:

- Datos Abiertos
- Interoperabilidad
- Visibilidad
- Preservación Digital
- Evaluación De Repositorios

UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ  
SISTEMA DE BIBLIOTECAS  
Biblioteca Virtual Universitaria  
San Luis Potosí, S.L.P.; México  
2017

ISBN 000-000-000-000-0  
180 p.

# MIEMBROS DEL COMITÉ



## Malgorzata Lisowska

Magister en Bibliotecología e información científica, de la Universidad Jagiellona de Cracovia - Polonia

Profesional con mas de 20 años de experiencia en gestión y procesos de cambio en las bibliotecas universitarias.

Desde 1999 - Directora del CRAI de la Universidad del Rosario en Bogotá, Colombia

2010-2013 - Coordinadora técnica del proyecto financiado por BID "LA Referencia"

2007-2010 - Miembro fundador de la Biblioteca Digital Colombiana



## Meilyn Garro Acon

Bachillerato en Ciencias de la Comunicación Colectiva con concentración en Producción Audiovisual de la Universidad de Costa Rica y Estudiante de la Maestría en Comunicación y Desarrollo de la Universidad de Costa Rica.

Coordinadora del Repositorio Institucional Kérwá desde el 2012 y miembro del equipo Latindex Costa Rica.

Líder público de Creative Commons Costa Rica y activista de la comunidad de Software Libre de la UCR.



## Caterina Marta Groposo Pavão

Universidade Federal do Rio Grande do Sul

País: Brasil

Bibliotecária do Centro de Processamento de dados da UFRGS desde 1994, doutora em Comunicação e Informação pelo Programa de Pós-Graduação em Comunicação e Informação (PPGCOM/UFRGS).



## Marisa Raquel De Giusti

<http://sedici.unlp.edu.ar>

PREBI-SEDICI, Universidad Nacional de La Plata Argentina

Biografía: Doctora en Ciencias Informáticas, Profesora en Letras e Ingeniera en Telecomunicaciones. Investigadora de la Comisión de Investigaciones

Científicas de la Provincia de Buenos Aires (CIC PBA), Directora de PREBI-SEDICI UNLP, y del Centro de Servicios en Gestión de Información (CESGI)



### **Rosalina Vázquez Tapia**

URL: <http://www.uaslp.mx>

Universidad Autónoma de San Luis Potosí  
(UASLP)  
México

Es Maestra en Tecnología Educativa por el Instituto Tecnológico de Estudios Superiores de Monterrey; Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de San Luis Potosí, y actualmente candidata a Doctora en Documentación y Tecnologías de Información en la Universidad Complutense de Madrid, España.

De 1991 a la fecha ha ocupado diferentes cargos en la Universidad Autónoma de San Luis Potosí (UASLP): Jefa del Centro de Información y Documentación Académica, de 1992 a 1999; Directora de Educación a Distancia, de 1999 al 2004; y del 2004 a la fecha como Directora de la Biblioteca Virtual Universitaria.

Es miembro fundador de la Red Abierta de Bibliotecas Digitales (RABID); Miembro fundador y Coordinador General de la Red Mexicana de Repositorios

Institucionales – REMERI; Responsable Técnico del nodo mexicano en la Red Latinoamericana de Repositorios Institucionales de Producción Científica LA-REFERENCIA; Socio técnico del proyecto Los Primeros Libros de las Américas. Además, desde el 2013 ha colaborado con el Consejo Nacional de Ciencia y Tecnología (CONACYT) como responsable técnico del proyecto “Estado del Arte y Diseño Metodológico para el desarrollo del

Repositorio Nacional de Producción Científica y Tecnológica”, y asesor en temas relacionados con la publicación



**Mag. Marlem Uribe Marengo**

Coordinadora Biblioteca Digital Uninorte,  
Colombia



# Indice

<b>MEMORIA DEL EVENTO</b> .....	2
<b>MIEMBROS DEL COMITÉ</b> .....	3
<b>PRESENTACIÓN</b> .....	7

## Conferencias magistrales..... 9

“Las dificultades de la preservación digital: problemas, desafíos y propuestas para los repositorios” .....	10
“Retos y oportunidades de la gestión de los datos abiertos de investigación” .....	11
“Google Scholar: redefiniendo el impacto académico” .....	12

## Mesas de ponencias ..... 13

<b>PRESERVACIÓN DIGITAL</b> .....	14
“La participación de México como miembro de la Biblioteca sobre el Patrimonio de la Biodiversidad (Biodiversity Heritage Library, BHL)” .....	15
“Un modelo de trabajo para agilizar la generación de documentos de texto para su preservación” .....	23
<b>INTEROPERABILIDAD</b> .....	37
“Propuesta de un esquema de metadatos, a partir del crecimiento del estándar de Dublin Core, para el Repositorio Universitario Especializado en Comunicación Pública de la Ciencia” .....	38
“PRÁTICAS DE INTEROPERABILIDADE PARA A INTEGRAÇÃO DE ACERVOS DIGITAIS HETEROGÊNEOS E DISTRIBUÍDOS” .....	47
“Desarrollo de un sistema de gestión de investigación (CRIS) para la consolidación del Repositorio Institucional NINIVE de la Universidad Autónoma de San Luis Potosí” .....	61
<b>VISIBILIDAD</b> .....	70
“Detección de bots en reportes estadísticos” .....	71
“Repositorios institucionales en Brasil” .....	78
“Módulo de Revisión por Pares en Abierto para Repositorios de Acceso Abierto” .....	91
“Programa de difusión de la producción intelectual del Instituto Tecnológico de Costa Rica.” .....	106
“Consolidación de la Red Mexicana de Repositorios Institucionales – REMERI e iniciativas de desarrollo y colaboración” .....	115
“Estrategia de Visibilidad de Recursos Educativos Abiertos para el Sistema de Educación Costarricense a través de un Repositorio Institucional” .....	122
“La producción académica de las Fuerzas Armadas Argentinas visible y abierta al mundo.” .....	131

## Panel ..... 140

<b>IMPLEMENTACIÓN DE ORCID EN REPOSITARIOS INSTITUCIONALES</b> .....	140
“Permitiendo la interoperabilidad de la información con identificadores” .....	141
“Estandarización de identificadores de investigadores con ORCID de la Universidad del Rosario de Colombia para la visibilidad y posicionamiento en web” .....	142
“Resultados de ORCID en la UAEMex dos años después” .....	143

## Talleres ..... 144

Taller 1: “El ciclo de vida de los datos de investigación desde la perspectiva de la ciencia en abierto” .....	145
Taller 2 “La gestión de la propiedad intelectual en el entorno digital” .....	146
Taller 3 “Conguración de directrices OpenAire 3.0 y uso de vocabularios controlados COAR en el marco europeo y latinoamericano” .....	147
Taller 4 “Cómo construir una identidad académica digital” .....	148

## Posters ..... 149

An implementation of thecnical Revision in DSpace Allowing Open Educational Resource Browser Access .....	150
Centro de Recursos para el Aprendizaje y la investigación .....	151
La nueva imagen del repositorio Académico Digital UANL y la implementación de Twitter para la difusión de contenidos .....	152
Unidad Académica en Estudios del Desarrollo, Universidad Autónoma de Zacatecas ..	153
Publicación de acceso abierto en revistas del núcleo central de la Web o Science en las áreas de biología y Química .....	154

<b>PATROCINADORES</b> .....	<b>155</b>
<b>FOTOGRAFÍAS DEL EVENTO</b> .....	<b>156</b>
<b>CONCLUSIONES</b> .....	<b>161</b>
<b>COMITÉ LOCAL</b> .....	<b>163</b>

# PRESENTACIÓN

La Universidad Autónoma de San Luis Potosí (UASLP), a través de la Biblioteca Virtual Universitaria del Sistema de Bibliotecas, y el Consorcio Iberoamericano para la Educación en Ciencia y Tecnología (ISTEC) con la colaboración de la Universidad del Rosario (UROSARIO), la Universidad del Norte (UNINORTE), la Universidad Nacional de la Plata (UNLP), la Universidad de Costa Rica (UCR), y la Universidad Federal de Río Grande del Sur (UFRGS), lo invitan a enviar sus propuestas de ponencia y poster para la Conferencia BIREDIAL-ISTEC, que se llevará a cabo del 17 al 19 de Octubre de 2016, en el Centro Cultural Universitario Bicentenario de la UASLP, San Luis Potosí; S.L.P.; México.



## **Esta sexta edición reúne los siguientes eventos:**

- VI Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'16)
- XI Simposio Internacional de Biblioteca Digitales (SIBD'16) de la Iniciativa LIBLINK del Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC)

La Conferencia BIREDIAL-ISTEC está dedicada a compartir el conocimiento de una amplia comunidad universitaria y de distintas disciplinas del saber, comprometida con la iniciativa de acceso abierto. El propósito del evento es exponer, discutir y analizar los temas que rodean a la iniciativa: materiales a disponer en acceso abierto, la problemática especial de los datos crudos de investigación, la disposición, preservación y difusión de las obras, la atención a los derechos de autor y la reproducibilidad de la investigación. Este año la Conferencia se dedica adicionalmente a tratar los nuevos tópicos del gobierno abierto y la evaluación y auditoría de los repositorios.

## **La Conferencia está dirigida a:**

- Comunidades científicas, observatorios, open data centers y otros.
- Profesionales interesados en temas relacionados con la gestión, promoción y visibilidad del conocimiento.

- Autoridades académicas.
- Investigadores, editores y directores de publicaciones científicas.
- Estudiantes de diversas especialidades.
- Instituciones de educación superior y centros de investigación.
- Personal del gobierno y la administración pública interesados en compartir inquietudes vinculadas a la transparencia de los datos públicos.

Ejes temáticos:

- **Datos abiertos:** Planes de gestión de datos (por ejemplo datos geoespaciales, estadísticos, datos médicos complejos: ADN, genoma, imágenes), reproducibilidad. Infraestructura: implementación de repositorios de datos, aplicaciones para visualización de datos primarios complejos. Gobierno abierto: gestión de datos abiertos en la administración pública.
- **Interoperabilidad:** Linked open data, directrices, protocolos, políticas y estándares de metadatos, redes de repositorios interoperabilidad de repositorios institucionales. El repositorio y la evaluación de la producción científica: interacción con sistemas de gestión de investigación (CRIS).
- **Visibilidad:** Políticas institucionales y públicas de acceso abierto, relación entre el acceso abierto y los indicadores de impacto de la producción científica, métricas alternativas, perfiles web, bibliometría, indicadores cienciométricos.
- **Preservación digital:** Planes de preservación digital y herramientas para generarlos, gestión de datos, curaduría de contenidos y vigilancia de formatos (automatización con herramientas de software en el ciclo de vida de los datos). Uso de metadatos de preservación y procesos de digitalización en repositorios (recomendaciones, normativas, estándares).
- **Evaluación de repositorios:** Criterios e indicadores de evaluación. Auditoría de repositorios institucionales (aplicación de la normativa ISO 16363). Ranking web de repositorios.

# Conferencias magistrales

# “Las dificultades de la preservación digital: problemas, desafíos y propuestas para los repositorios”

*Dra. Marisa R. De Giusti  
Universidad Nacional de la Plata, Argentina*



Dra. Marisa R. De Giusti, Universidad Nacional de la Plata, Argentina  
Conferencia: “Las dificultades de la preservación digital: problemas, desafíos y propuestas para los repositorios”  
Marisa R. De Giusti es Ingeniero en Telecomunicaciones, Profesora en Letras y Doctora en Ciencias Informáticas, títulos todos otorgados por la Universidad Nacional de La Plata. Se desempeña como Profesora en el Grado y en el Doctorado de la Facultad de Informática de la UNLP y es Investigadora Principal de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CICBA).  
Directora del Centro Propio de CICBA denominado Centro de Servicios en Gestión de Información (CESGI).  
Directora de los Programas de la UNLP PREBI-SEDICI y de los Repositorios Institucionales SEDICI, CIC Digital y el Repositorio de Datos OMLP (CIC-Conicet-UNLP).  
Miembro del Comité de Expertos del Sistema Nacional de Repositorios Digitales del MinCyT (Argentina)  
Miembro de la Junta de Directores del Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC) y Directora de la Iniciativa LIBLINK (Library Linkage) de ISTEC.  
Sus temas de investigación están centrados en los repositorios institucionales con un fuerte acento en la preservación de los objetos digitales y la evaluación de los repositorios.

# “Retos y oportunidades de la gestión de los datos abiertos de investigación”

*Dra. Remedios Melero  
CSIC, España.*



Dra. Remedios Melero, Universidad de Valencia, España

Conferencia: “Retos y oportunidades de la gestión de los datos abiertos de investigación”

Doctora en Ciencias Químicas, por la Universidad de Valencia. Trabaja en el Instituto de Agroquímica y Tecnología de los Alimentos del Consejo Superior de Investigaciones Científicas de España (CSIC) como científico titular y es editora de la revista *Food Science and Technology International*. Vicepresidenta de European Association of Science Editors hasta el año 2012. Es profesora en el master de Documentación Digital impartido por la Universidad Pompeu Fabra y miembro del comité científico de Redalyc y de Scielo España. Sus temas de trabajo incluyen publicaciones electrónicas, procesos de arbitraje académico, acceso abierto a la producción científica y propiedad intelectual. Es miembro del grupo de trabajo Acceso Abierto a la Ciencia para el estudio de revistas, repositorios y políticas institucionales de acceso abierto. Ha participado en el proyecto europeo NECOBELAC para la difusión y creación de redes en torno a la publicación científica y el acceso abierto a la ciencia, entre Latinoamérica y Europa. Actualmente trabaja en un proyecto nacional coordinado con la Universitat de Barcelona sobre el grado de implantación y sostenibilidad de recursos en acceso abierto, y hasta julio de 2016 ha participado en el proyecto europeo FOSTER (Facilitate Open Science Training for European Research).

# “Google Scholar: redefiniendo el impacto académico”

*DR. Enrique Orduña-Malea  
Universidad Politécnica de Valencia (UPV)*



DR. Enrique Orduña-Malea, Universidad Politécnica de Valencia (UPV)

Conferencia: “Google Scholar: redefiniendo el impacto académico”

Es Ingeniero Técnico de Telecomunicaciones, Licenciado en Documentación, Master en contenidos multicanal y Doctor en Documentación (Tesis doctoral orientada a la cibermetría académica) por la Universidad Politécnica de Valencia (UPV).

Es autor de más de 50 artículos y capítulos de libro académicos y profesionales, entre las que se incluyen más de 25 publicaciones recogidas en los Journal Citation Reports (JCR). Es co-autor de las monografías “Cibermetría: midiendo el espacio red” y “La revolución Google Scholar”, de inminente publicación.

Desde 2008 es miembro del Grupo ThinkEPI (Grupo de Análisis sobre Estrategia y Prospectiva de la Información); desde 2012 del Grupo de investigación EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada (UGR); y, desde 2014 es miembro fundador del Grupo Trademetrics, orientado a la aplicación de indicadores web en el análisis de la presencia e impacto de las empresas y marcas en la Web.





# Mesas de ponencias



# Preservación Digital

Moderador: Caterina G. Pavão (UFRGS)

# “La participación de México como miembro de la Biblioteca sobre el Patrimonio de la Biodeversidad (Biodiversity Heritage Library, BHL)”

*Lic. María del Carmen Ramírez González  
Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, México.  
correo electrónico: mramirez@conabio.gob.mx*

---

## **Resumen**

La Biblioteca sobre el Patrimonio de la Biodiversidad (Biodiversity Heritage Library, BHL) surgió 2007 por iniciativa de museos y bibliotecas de Estados Unidos e Inglaterra con el apoyo de Internet Archive. BHL funciona a manera de consorcio en el que trabajan bibliotecas de instituciones botánicas y de historia natural para ofrecer libre acceso en línea a sus colecciones dirigido a biólogos, taxónomos, ecólogos, bibliotecarios, estudiantes y público en general, para quienes pone a disposición un acervo de más de 170 mil volúmenes que incluyen más de 46 millones de páginas.

En el 2014 la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) fue invitada a ser miembro de BHL para lo cual se decidió conformar un nodo para México e invitar a participar a las instituciones nacionales especializadas en biodiversidad que además aportaran materiales en español; actualmente BHL México cuenta con la colaboración de seis instituciones mexicanas, aunque se espera que otras más, que han aportado mucho al conocimiento de la biodiversidad mexicana, se sumen a este proyecto.

El objetivo de la ponencia, es dar a conocer la experiencia de México y el flujo de trabajo para integrar las obras resguardadas en las bibliotecas e instituciones que participan y así, animar a otras instituciones a participar ya que aún falta mucho para consolidar BHL México con ayuda y la participación entusiasta de los bibliotecarios y de un equipo multidisciplinario se logrará contar con una excelente colección de obras que permita contribuir a enfrentar los retos del conocimiento de la riqueza biológica del país y su difusión.

## **Abstract:**

The Biodiversity Heritage library emerged in 2007 through the initiative of some museums and libraries in the USA and Britain and with the support of the Internet Archive. BHL works as a consortium in which libraries of botanical and natural history institutions work together to offer open access to their collections and is aimed at biologists, taxonomists, ecologists, librarians, students and the general public, who have access to more than 170 thousand volumes that include over 46 million pages.

In 2014 the National Commission for the Knowledge and use of Biodiversity (CONABIO) was invited to be a BHL member and it was decided that it would be a node for Mexico and would invite national institutions specialized in biodiversity who could also provide materials in Spanish; currently, BHL Mexico has the support of six Mexican institutions, although it is expected that others, who have contributed much to the knowledge of the Mexican biodiversity will join this project.

The aim of this talk is to present the experience of Mexico and the workflow to integrate the items from libraries and institutions that participate in this project and encourage other institutions to participate to consolidate BHL Mexico, with the support and enthusiastic participation of librarians and a multidisciplinary team, we will achieve an excellent collection of works that will assist in meeting the challenges of the knowledge of the biological wealth of the country and its dissemination

**Palabras clave:** Biblioteca digital, biodiversidad, metadatos, flujo de trabajo

**Key words:** Digital Library, biodiversity, metadata, workflow

La Biblioteca sobre el Patrimonio de la Biodiversidad (Biodiversity Heritage Library, BHL) surgió en 2007 por iniciativa de museos y bibliotecas de Estados Unidos e Inglaterra, con ayuda de otros proyectos como Internet Archive (IA) y Encyclopedia of Life (EOL) se conformó un consorcio en el que trabajan bibliotecas de instituciones botánicas y de historia natural, con el objetivo de ofrecer libre acceso en línea a sus colecciones a biólogos, taxónomos, ecólogos, bibliotecarios, estudiantes y público en general.

A partir de 2009 se sumaron a este proyecto socios de varios países y regiones estableciendo nodos de BHL en Europa, China, Australia, Brasil, Egipto, África, Singapur, Canadá y México. Sus principios son el acceso abierto, la colaboración, la interoperabilidad y la transparencia y con estos busca el acceso a la literatura sobre biodiversidad, principalmente aquella con alto contenido de nombres científicos, la repatriación de la literatura, la reutilización de las obras de dominio público y el trabajo cercano con editores no comerciales. Se trabaja en una estructura automática de inteligencia taxonómica, identificadores digitales y digitalización de grandes cantidades de documentos.

El equipo de trabajo es muy pequeño, se compone por cuatro personas dedicadas en tiempo completo, cuatro directivo que también trabajan para otras instituciones y finalmente el personal denominado staff que, sumando las horas dedicadas a BHL, se cuenta con aproximadamente el equivalente al trabajo de 17 personas de tiempo completo, entre ellos bibliotecarios, tecnólogos, restauradores, catalogadores, taxónomos, administradores, algunos voluntarios y estudiantes en prácticas profesionales.

Para que este trabajo se pueda llevar a cabo de manera organizada y respetando es-

tándares internacionales, BHL pone a disposición de todos sus colaboradores documentación para cada proceso, sus miembros y afiliados (organizados en comités) elaboran y actualizan documentación exhaustiva, lo que permite a las instituciones evitar duplicados, obtener imágenes de alta calidad apegados a estándares internacionales para la digitalización, la uniformidad en la asignación de metadatos a nivel de página y la organización de las obras. Un documento de los más importantes es la política de desarrollo de colecciones donde se describe el flujo de trabajo, mismo que se divide en cinco grandes pasos:

## **1) Selección**

Cada biblioteca o institución es responsable de llevar a cabo una selección adecuada de los títulos que conformarán su colección, que respondan a la Política de Desarrollo de Colecciones de BHL y al mismo tiempo atiendan las necesidades de información de sus usuarios. En esta etapa se deberán considerar en primer lugar aquellas obras raras, ejemplares únicos y publicaciones de la propia institución. Además BHL publica un listado de títulos a los que se han asignado prioridades de digitalización acorde con las solicitudes de los propios usuarios de la plataforma. Un paso muy importante de la selección es la negociación de los permisos o bien el estatus de derechos de autor.

La colección está integrada por libros, revistas y libretas de campo principalmente de dominio público, aunque también se han negociado una gran cantidad de permisos sobre obras de reciente publicación, lo que sin duda contribuye en gran medida a la democratización de la ciencia y representa un gran esfuerzo por parte de cada biblioteca en materia de derechos de autor. Al respecto, BHL opera bajo una política de reconocimiento de las instituciones y bibliotecas que contribuyen, pero cada institución es responsable de realizar la investigación del estatus de los derechos de autor antes de iniciar el proceso de digitalización y respetar las leyes de su país.

Una vez que se ha decidido por los títulos que se digitalizarán se deberá revisar exhaustivamente el catálogo de BHL para evitar duplicar esfuerzos y recursos en la digitalización de un ítem que ya se encuentra en línea. Pudiendo suceder que el ejemplar tiene alguna peculiaridad y por ello se decide digitalizar y subir.

## **2) Digitalización**

Obtenido el listado de obras, se procede a digitalizar de cubierta a cubierta, incluyendo las hojas en blanco, con los estándares establecidos para las impresiones en blanco y negro (mínimo 300 ppi) y a color (mínimo 600 ppi). Este proceso se compone de varios pasos que incluyen el manejo de un software gratuito para los miembros de BHL que permite asignar metadatos sobre los derechos de autor, la descripción del material (para lo que se requiere una ficha catalográfica en formato MARCXML); metadatos para la identificación de ilustraciones, mapas, portadas, entre otros. Este software llamado Macaw envía los archivos ya procesados a Internet Archive, actualmente es la única forma de enviar los archivos y compartirlos.

## **3) Usuarios**

Los usuarios forman parte de este ciclo ya que con la retroalimentación que proporcio-

nan, por medio de la página, es posible realizar correcciones y ayudar a realizar mejoras a la colección y sus servicios.

#### 4) Curación

Por medio un tablero de administración se realizan las correcciones necesarias en la ficha bibliográfica, corrección de imágenes, orden de ejemplares, entre otros. Es importante invertir tiempo en realizar las correcciones pertinentes para mejorar la calidad de los metadatos y facilitar la recuperación de los ítems.

Concluido este ciclo, los usuarios pueden consultar las obras desde varias plataformas de acceso abierto como son: BHL (biodiversitylibrary.org), Internet Archive (<https://archive.org>) en la colección de BHL, en la Digital Public Library of America (DPLA- <https://dp.la/>) con los ítems inmersos en su propia colección y BioStor (<https://biostor.org>) donde se pueden localizar los artículos de las obras periódicas de la colección de BHL; en Encyclopedia of Life (<https://eol.org>) en cada ficha de la especie descrita se encuentra la literatura citada y links a las obras de BHL en las que se refiere la especie, de la misma manera sucede en Enciclovida ([www.enciclovida.mx](http://www.enciclovida.mx)) y en Flickr ([www.flickr.com/photos/biodivlibrary](http://www.flickr.com/photos/biodivlibrary)) se pueden visualizar las ilustraciones de las obras de dominio público.

Internet Archive tiene una función muy importante que es almacenar las imágenes y archivos de los libros escaneados que después se visualizarán en BHL, por lo que desde su página se contabilizan las visitas por ítem; en el caso de Encyclopedia of Life tienen una función recíproca para intercambio de nombres de las especies, con ayuda de la minería de textos y el Optial Character Recognition (OCR) realiza el reconocimiento de los nombres científicos en cada página con un algoritmo llamado Global Names Recognition and Discovery (GNRD) y al mismo tiempo se obtienen nuevos nombres. De allí la importancia de obtener imágenes de gran calidad, ya que el proceso de OCR no sería posible si cada letra no se reconoce correctamente y por lo tanto el programa que reconoce los nombres científicos de las especies también fallará, perdiendo información valiosa para los usuarios. Finalmente, pero no menos importante, a cada título se le asigna un Digital Object Identifier (DOI) en caso de ser necesario.

A continuación se presentan algunas estadísticas generales sobre BHL que dan cuenta del esfuerzo y la colaboración multidisciplinaria que implica el acceso abierto:

<b>títulos</b>	109 515	nombres científicos	168 509 647
<b>volúmenes</b>	184 467	DOI	103397
<b>páginas</b>	498 621 61	Tweeter (seguidores)	7355
<b>Flickr</b>	104 680	Facebook (seguidores)	12911
<b>Segmentos</b>	190 255	Flickr (seguidores)	8158

**Tabla 1. Estadísticas de BHL Central**

Desde el 2010, BHL ha recibido seis reconocimientos internacionales por su excelente labor sobre el acceso abierto, el trabajo en equipo y el desarrollo de herramientas innovadoras que destacan la colaboración internacional y el mérito al servicio para el estudio de la Botánica:

<i>Fecha</i>	<i>Reconocimiento</i>	<i>Descripción</i>
2010	John Thackray medal	Premio entregado a BHL en reconocimiento por los últimos tres años conformando la biblioteca digital sobre biodiversidad.
2010	Outstanding Collaboration award. Association for Library Collections & Technical Services (ALCTS)	Por su notable trabajo colaborativo con museos, bibliotecas e instituciones de investigación, organizados para digitalizar preservar y libre acceso a literatura sobre biodiversidad.
2012	BHL - Australia) Victorian Government Arts Leadership Recognition Award	Como reconocimiento al trabajo del voluntariado para la colección de BHL Australia.
2013	Laureate IDG's Computerworld Honors Program	
		Premio anual de honor entregado por la visión sobre el uso de las tecnologías de información que promueven un cambio social, económico y educativo.
2013	Charles Robert Long Award of Extraordinary Merit from the Council on Botanical and Horticultural Libraries (CBHL).	
		Por la colaboración a gran escala de las bibliotecas sobre biodiversidad poniendo al descubierto colecciones a los estudiantes y científicos.

<i>Fecha</i>	<i>Reconocimiento</i>	<i>Descripción</i>
2015	Internet Archive's Internet Heroes award	Como reconocimiento por la digitalización de la herencia cultural.

**Tabla 2. Reconocimientos otorgados a BHL Central**

Sin duda, el reconocimiento más grande es aquel que los usuarios otorgan a BHL por sus colecciones y el servicio de referencia. Constantemente se reciben comentarios de usuarios que destacan la importancia de las colecciones en sus investigaciones y sobre todo el libre acceso a obras antiguas (libros, publicaciones periódicas y libretas de campo principalmente que datan del siglo XV a la fecha); esto se traduce en más de 1.5 millones de visitas al año y 80 millones de descargas de ítems. La Ciudad de México se encuentra en el top diez de ciudades que más consultas realizan, después de Londres, París, Nueva York y Washington.

Una vez que los materiales se visualizan en cualquier portal sea BHL, IA u otro proyecto en el que se participe, cada ítem tiene una licencia de Creative Commons, que permite modificar, compartir y reutilizar las obras siempre y cuando no sea para uso comercial. Sin embargo, el trabajo no termina cuando la obra se visualiza correctamente en el portal de BHL, ahora es necesario dar visibilidad a estos títulos que llevó tiempo y dedicación colocar en la colección, esto se hace por medio de las redes sociales, el libro del mes, las campañas por colecciones especiales, la colección de imágenes en Flickr. Es un proceso cíclico que toda biblioteca conoce perfectamente.

Con una idea somera sobre este proceso, en diciembre de 2014 la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) fue invitada a ser miembro de BHL y aceptaron firmando el Memorandum of Understanding y se convirtieron en el miembro número 17 de BHL, se aceptaron los términos de participación en cuanto a las políticas de desarrollo de colecciones así como los estándares de calidad tanto para los metadatos como para las imágenes y archivos que se integrarían.

En ese mismo momento se decidió conformar un nodo para México, en lugar de participar como una sola institución, y se extendió la invitación a instituciones nacionales especializadas en biodiversidad que además tuvieran la oportunidad de aportar materiales en español. Este nodo global se nombró BHL México y se diseñó un logotipo.

En abril de 2015 se invitó a los directivos de doce Instituciones de Investigación para la Digitalización de Literatura sobre Biodiversidad de México quienes firmaron una carta de intención de participación donde reconocían la importancia de la difusión de la producción científica y de divulgación con ayuda de BHL. A finales del mismo año se les solicitó nombraran un representante que asistiría a un taller técnico y formaría parte del personal staff de BHL México; con una lista preliminar de 15 posibles asistentes se iniciaron los preparativos.

El taller se llevó a cabo del 2 al 4 de diciembre de 2015 y se denominó “Primer Taller Técnico de BHL México”, participaron bibliotecarios, investigadores e ingenieros en sistemas, en las instalaciones de CONABIO y asistieron 17 representantes de 9 instituciones. Durante cerca de 24 horas, repartidas en tres días de trabajo intensivo, el personal de BHL Central se dio a la tarea de presentar el flujo de trabajo, atender las dudas de los



asistentes y hacer demostraciones sobre el uso del software, así como los canales de comunicación para el personal Staff que dedica tiempo para el desarrollo y mejora de la colección.

Como resultado de este taller se acordó tener reuniones subsecuentes para afinar detalles con algunos directivos, elaborar un plan de trabajo sujeto a aprobación de los asistentes, adecuar la política de desarrollo de colecciones y finalmente coincidieron que era necesario una comunicación constante para discutir temas de interés sobre los materiales que integrarían y ser un grupo de apoyo para atender dudas, principalmente sobre derechos de autor.

A partir de enero de 2016 se envió, a los asistentes al taller, documentación en español relevante para apoyar la próxima integración de materiales, la guía de uso del software, políticas, estándares de calidad de imágenes y metadatos y la política de desarrollo de colecciones con las observaciones pertinentes.

Durante la primera mitad del año se atendieron reuniones de capacitación con los representantes de bibliotecas que se encuentran en la Ciudad de México y con aquellas personas en provincia se atendieron dudas por teléfono y por Skype. Además de tres teleconferencias grupales para atender dudas y dar seguimiento a la participación de las instituciones.

Esto dio como resultado reunir 623 ítems con la colaboración de seis instituciones:

Institución	Títulos	Volúmenes	Páginas	Descargas	Redes sociales
Biblioteca Conjunta de Ciencias de la Tierra- ICMYL, UNAM	19	19	3 051	666	15 tweets 1 mención en Facebook
Biblioteca, Instituto de Biología, UNAM	3	71	3 636	1 226	0
CONABIO	22	158	15 819	4 729	68 tweets 3 menciones en Facebook
Instituto de Ecología A.C.	4	337	20 690	8 227	0
Instituto de Ecología, UNAM	1	36	470	715	0
El colegio de la Frontera Sur	1	2	88	27	0
<b>TOTAL</b>	<b>50</b>	<b>623</b>	<b>43 754</b>	<b>15 590</b>	<b>87</b>

**Tabla 3. Inventario y estadísticas de uso de BHL México**

Dos instituciones más se encuentran en proceso de integrar ítems y en mayo de 2016 se enviaron cinco invitaciones más, ya que aún falta mucho para consolidar BHL México como el referente de biblioteca digital sobre biodiversidad. Aún de deben integrar gran cantidad de materiales, el desarrollo de un portal para consulta web en español, promocionar la colección, apoyar el proceso de mejora continua en el catálogo y otros

proyectos de minería de datos para el reconocimiento de nombres científicos y comunes de las especies. Será gracias a la participación de las instituciones, de sus bibliotecarios y de un equipo multidisciplinario que se logrará contar con una excelente colección de obras de importancia histórica y actual que permita contribuir a enfrentar los retos del conocimiento de la riqueza biológica del país y su difusión.

### **Bibliografía consultada:**

Biodiversity Heritage Library: BHL featured in Digital Public Library of America Hackathon. (2012). Blog.biodiversitylibrary.org. Retrieved 2 August 2016, from <http://blog.biodiversitylibrary.org/2012/04/bhl-featured-in-digital-public-library.html>

Biodiversity Heritage Library: BHL Welcomes CONABIO as newest Member and Global Node!. (2015). Blog.biodiversitylibrary.org. Retrieved 2 August 2016, from <http://blog.biodiversitylibrary.org/2015/01/bhl-welcomes-conabio-as-newest-member.html>

Gwinn, N. E., & Rinaldo, C. A. (2008). The Biodiversity Heritage Library: sharing biodiversity literature with the world. IFLA Conference Proceedings, 1-23.

Koleff Osorio, P. & Ramírez González, M. (2016). México, miembro de la Biblioteca Digital Internacional sobre Biodiversidad. Noticiero De La AMBAC, 87(enero-marzo), 14-17.

Pilsk, S. C., Person, M. A., Deveer, J. M., Furfey, J. F., y Kalfatovic, M. R. (2010). The Biodiversity Heritage Library: Advancing Metadata Practices in a Collaborative Digital Library. Journal Of Library Metadata, 10(2/3), 136-155. doi:10.1080/19386389.2010.506400

Randall, P. (2016). Purposeful Gaming and the Biodiversity Heritage Library. Journal Of Agricultural & Food Information, 17(1), 71-76. doi:10.1080/10496505.2016.1116339

Smith, J. E., y Rinaldo, C. A. (2015). Collaborating on open science: The journey of the Biodiversity Heritage Library. Information Services & Use, 35(4), 211-216. doi:10.3233/ISU-150785

# “Un modelo de trabajo para agilizar la generación de documentos de texto para su preservación”

*Salamone Lacunza, Paula ; Villarreal, Gonzalo L. ; De Giusti, Marisa R. ; Lira, Ariel J .*

---

## Resumen

Los repositorios institucionales (RI) tienen la responsabilidad de gestionar, preservar y ofrecer acceso libre a la producción científica de una institución particular. Para ello, el repositorio debe establecer políticas que aseguren la autenticidad de los objetos digitales, que prevengan la pérdida parcial o total de los mismos, y que permitan acceder a su contenido por una comunidad de usuarios designada. Para la correcta ejecución de estas políticas se deben realizar un conjunto de actividades de preservación que, idealmente, deberán integrarse al conjunto de tareas de administración del repositorio cotidianas, y así estandarizar y asegurar la realización de las actividades de preservación que se hacen sobre los objetos digitales. Desde luego, también será necesario realizar revisiones periódicas sobre los métodos y circuitos implementados, estudiar la efectividad de las herramientas y formatos en uso, y realizar perfilamientos y análisis de los objetos digitales del repositorio a fin de controlar la eficacia de las tareas de preservación. Como es de suponerse, las actividades de preservación pueden requerir una importante carga adicional para los administradores del repositorio. Por ejemplo, como se mencionó, será necesaria la incorporación manual o la verificación y corrección, si son incorporados por el software, de un nuevo conjunto de metadatos de preservación que pueden ser descriptivos (soporte, identificadores), estructurales (capítulos, índices, relaciones) y administrativos (formato, versión del software, resolución, compresión).

Además de la inclusión de los metadatos, la preservación digital implica el análisis de los formatos de los archivos digitales que se ingestan al repositorio, la selección del mejor formato de transformación o migración, la transformación o migración en sí desde el archivo original hacia su correspondiente formato preservable, la verificación de la correcta transformación a fin de comprobar que no se han generado efectos indeseados que impidan la reproducción apropiada del contenido, la validación según las reglas del estándar utilizado y el cumplimiento de las normativas requeridas para el archivo resultante según el formato al cual ha sido migrado, y finalmente su almacenamiento en un medio adecuado. Sin embargo, las actividades de preservación no finalizan aquí: los archivos almacenados deben almacenarse en distintos medios mediante un sistema de

copias de seguridad desatendidas (sin intervención humana), en lo posible geográficamente distribuidas, a fin de asegurar que no se perderán en caso de catástrofes; también deberán realizarse controles periódicos a fin de asegurar su integridad, y es aconsejable una revisión periódica de los formatos para asegurar la mejor selección de los mismos atendiendo los supuestos de la preservación digital. En fin, no basta con establecer un plan de preservación, sino que es necesario ejecutarlo y revisarlo periódicamente.

## **Materiales y metodología**

En este trabajo se hará énfasis en el conjunto de tareas que realiza la administración del repositorio relativas al análisis, transformación y validación de los objetos digitales a los fines de la preservación. También se expondrán algunos casos especiales detectados, en donde la conversión del objeto digital no sigue un camino estándar establecido y que por lo tanto requieren una evaluación individual para determinar su tratamiento. Se mencionarán las pruebas y análisis realizados con un conjunto de herramientas informáticas y utilizando objetos reales del repositorio SEDICI de la UNLP.

## **Resultados parciales y conclusiones**

Se presentará aquí un modelo de trabajo semi-automático, mediante el cual los administradores delegan el análisis y transformación de estos objetos a un conjunto de herramientas informáticas, las que a su vez brindan un reporte de las tareas realizadas y los resultados obtenidos. Del conjunto de herramientas analizadas, se destacarán aquellas que fueron finalmente seleccionadas para la realización de las actividades de preservación y los motivos de su selección. También se explicará en detalle la metodología de trabajo implementada a fin de agilizar las tareas de la administración del repositorio a la hora de transformar los objetos digitales y disminuir la carga de procesamiento de los equipos informáticos.

El modelo aquí propuesto se ha implementado por el momento en los documentos de texto, para los que se utiliza el formato PDF/A (en alguna de sus variantes) descrito en las normas ISO 19005-1, ISO 19005-2 e ISO 19005-3, pero como se verá, su diseño e implementación permite fácilmente la incorporación de otros tipos de documentos no textuales, como por ejemplo imágenes o archivos de audio.

**Palabras claves: repositorios institucionales, preservación digital, metadatos de preservación, actividades de preservación.**

## **Abstract**

Institutional repositories have the responsibility to manage, preserve and give access to the scientific production of the institution. To this end, they must establish policies that ensure digital objects authenticity, that avoid partial or total lost of files, and that make contents available to the designated users' community. These policies must be carried out through a set of preservation activities that, preferably, should be integrated into the everyday administration tasks. Clearly, periodical reviews about the methods, tools and workflows under use must be performed, as well as profilings over random samples from the repository, in order to make sure that the preservation plan is being kept and that results are correct according to the current formats and access platforms.

Preservation activities, such as verification and correction of metadata, impose an addi-

tional workload to the repository administration. Besides these efforts, correct documents preservation require to understand file formats, and to transform files into preservable versions if possible, which in turn require also to check for correctness of transformed objects to make sure they can still be opened and accessed. In addition to this work, data preservation requires every chunk of information to be mirrored in more than one device, preferably in geographically distributed environments. Ideally, this must be done transparently for users and on an automatic basis.

This work is focused on all the administration tasks required to analyze, transform and validate digital objects to preserve them. A semi-automatic work model is presented here, in which administrators delegate the analysis and transformation of these objects to a set of computing tools, which besides the actual execution of tasks, give back to the administrators a report of all performed events and its results. It will also explained in detail the workflow performed by the staff of the repository to speed up all tasks and to minimize the processing load of computing hardware.

The model proposed here has been implemented for text documents so far. PDF/A format and its variants (ISO 19005-1, ISO 19005-2 and ISO 19005-3) has been used here. However, this implementation allow the addition of other file types, such as audio or images.

**Keywords: institutional repositories, digital preservation, metadata preservation, preservation activity.**

## **Introducción**

La preservación de la producción científica y académica de las instituciones cobra mayor relevancia a medida que las mismas generan recursos, resultantes de los conocimientos de las personas y de sus expresiones, que nacen cada vez más en formas digitales (De Giusti et al. 2014; De Giusti et al. 2012). Los productos de origen digital son valiosos, constituyen un verdadero patrimonio a conservar a futuro para la sociedad, pero pueden no contar con un respaldo físico (por ejemplo en papel). Cabe destacar que el simple hecho de almacenar copias digitales de estos recursos no es suficiente para preservarlos a lo largo del tiempo: los objetos digitales (OD) deben ser descritos correctamente para, por ejemplo, conocer su trayectoria y modificaciones, deben utilizarse formatos apropiados que permitan continuar transformándolos a medida que surgen nuevos formatos o plataformas de acceso, deben almacenarse de forma tal de evitar las pérdidas por fallas de los dispositivos de almacenamiento, etcétera (PREMIS 2016; CCSDS 2002). Estos son sólo algunos de los desafíos que las instituciones deben enfrentar a la hora de asegurar la perpetuidad de su patrimonio digital, y es aquí donde los repositorios digitales cobran un rol central gracias a su posicionamiento en la organización como principal medio para almacenar, dar visibilidad y difundir la producción institucional.

## **Contexto**

Este proyecto se realiza en el marco del repositorio SEDICI, perteneciente a la Universidad Nacional de La Plata (De Giusti et al., 2008). El repositorio alberga recursos de todas las unidades académicas de la UNLP, así como también de sus programas, direcciones y proyectos. Esto constituye una de las características principales de SEDICI: un flujo constante de recursos ingresan constantemente al repositorio, de orígenes y tipologías muy variadas, que son verificados, catalogados, adaptados, preservados y difundidos



por un conjunto de personas (administradores del repositorio). Estos administradores pueden cumplir distintos horarios laborales, poseen distinto grado de conocimientos técnicos, y sus entornos de trabajo pueden alterarse a lo largo del tiempo: compra de nuevas computadoras, instalación de nuevos sistemas operativos, cambios en las aplicaciones que utilizan (navegadores, herramientas de oficina, sistemas de seguridad), etcétera. Si bien esta variabilidad otorga una gran flexibilidad desde el punto de vista organizacional, puede presentar un desafío a la hora de disponer de las herramientas que aseguren el correcto y equitativo procesamiento de los OD que se alojan en el repositorio. Una de las ventajas principales del modelo que se propone en este trabajo es la separación entre los equipos de trabajo de los administradores y el software de procesamiento de documentos, lo cual ayuda a mantener la flexibilidad de este tipo de entornos. Por otro lado, como se verá más adelante, el modelo es fácilmente escalable y replicable, lo que constituye una enorme ventaja tanto si la carga de trabajo continúa creciendo como también para otras organizaciones que deseen implementar un modelo similar.

Al día de hoy SEDICI cuenta con cerca de 50 mil recursos, y el 95% de ellos son documentos de texto en formato PDF (De Giusti, 2014). Si bien el modelo aquí propuesto no está sujeto al formato PDF exclusivamente, debido al elevado porcentaje de este formato en su primera implementación para este repositorio, se trabajó sobre documentos de texto en este formato.

### **Formatos de preservación**

La preservación de objetos digitales requiere, como se comentó arriba, el uso de formatos apropiados que aseguren el acceso a los datos alojados dentro de dichos objetos y que permitan realizar transformaciones y adaptaciones para asegurar su acceso a medida que las herramientas de acceso avanzan y que surgen nuevos formatos de archivos. Esto se aplica a cualquier tipo de recurso digital, pero como es de esperarse, la selección del formato de preservación deberá adecuarse al tipo concreto de OD: documentos de texto, imágenes, audios, videos, datos crudos, objetos espaciales, etcétera (Kresse et al. 2015). A la hora de seleccionar un formato de preservación, deben tomarse en consideración diversos aspectos, como por ejemplo el grado de estandarización del formato, el nivel de apertura de la licencia de dicho formato, la posibilidad de pérdida de información al transformar OD hacia ese formato (por ejemplo, al utilizar algoritmos de compresión con pérdida), la facilidad a la hora de transformar OD hacia estos formatos, el tamaño final de los archivos, entre otros (Brown, 2008; Giménez Chornet, 2014).

En el caso particular de los documentos de texto, el estándar PDF/A, descrito en las normas ISO 19005-1, ISO 19005-2 e ISO 19005-3, se impone como el formato más apropiado para su preservación. Este formato está basado en el estándar PDF 1.4, al que le incorpora algunos requerimientos adicionales, como ser:

- 1.** Especificaciones sobre los metadatos y la estructura del archivo.
- 2.** La paleta de colores (incluyendo escala de grises y blanco/negro) no deben ser representados en un espacio de color de dispositivo (DeviceRGB, DeviceCMYK, DeviceGray).
- 3.** Las fuentes usadas en texto visibles deben estar embebidas (incluidas dentro del archivo).
- 4.** Para un PDF/A-1a, la estructura original del documento se mantendrá igual al

documento PDF/A original. Es decir, no se crearán nuevas etiquetas y la estructura no deberá cambiar, por lo que para crear un archivo PDF/A-1a, el OD debió haber sido creado estructurado y etiquetado. En cualquier otro caso, se optara por PDF/A-1b.

La principal diferencia entre PDF/A-1b y PDF/A-1a, es que PDF/A-1a posee especificaciones adicionales sobre PDF/A-1b, como por ejemplo:

1. Las fuentes embebidas tienen otros requerimientos, como ser, su representación en Unicode (ISO 19005-1, capítulo 6.3.8).
2. El documento debe contener una estructura lógica (ISO 19005-1, capítulo 6.8).

Uno de los propósitos de los requerimientos del estándar PDF/A-1a es de proveer soporte para personas con discapacidades diferentes, por ejemplo, incorporando la información requerida y necesaria para aplicaciones que hagan el pasaje de texto a voz. La estructura lógica del documento es una descripción del contenido de las páginas, que debe ser dada por el documento original y consiste en un correcto etiquetado jerárquico que distingue el verdadero contenido de los artefactos incluidos en el documento (números de páginas, pies de página, artefactos de maquetación, etc.). El etiquetado provee una breve descripción, y esta debe ser sencilla y fácilmente comprensible para las personas, por eso no puede ser generada en una conversión (no es posible derivarla automáticamente por el software de conversión) sino que debe ser generada al momento de la creación del OD. Esta es una de las razones por la cual no cualquier PDF puede ser convertido a PDF/A-1a.

El PDF/A-2 descrito en la norma ISO 19005-2, está basado sobre la norma ISO 32000-1, correspondiente al estándar PDF 1.7, y este a su vez es una extensión complementaria del estándar PDF/A-1. Las principales diferencias entre PDF/A-1 y PDF/A-2 son:

1. Se agrega la compresión de JPEG2000 para imágenes.
2. Las imágenes con transparencias son permitidas.
3. Contenidos adicionales (conocidos como layers o capas) pueden ser visibles o no.
4. Muchos PDF/A pueden combinarse para crear un solo archivo.
5. Se crea un nuevo nivel de conformidad, el nivel U (unicode), que permite crear un PDF de búsqueda sin tener que cumplir con el estricto nivel de conformidad A (A por accesibilidad).

Los documentos que contengan las características descritas anteriormente, en particular las transparencias o layers, deberán ser convertidos a PDF/A-2, rara vez a PDF/A-1. El estándar PDF/A-3, descrito en la norma ISO 19005-3, está basado sobre ISO 32000-1 (PDF 1.7). Es a su vez una extensión complementaria de PDF/A-2. Algunas de las diferencias principales entre PDF/A-2 y PDF/A-3 son:

1. Los archivos, de cualquier formato y conformancia, deben ser embebidos en el PDF/A creado.
2. Los archivos embebidos pueden ser asociados a cualquier parte del PDF/A-3 creado.

A partir de las descripciones previas, es evidente que sería ideal alcanzar el estándar PDF/A-1a. Sin embargo, esto no siempre es posible debido a las características particulares de cada archivo. En caso de no poder llegar a ese formato, se aceptan también otros formatos de PDF/A.

### **Propuesta de servidor de validación y transformación**

La realización de actividades de preservación puede requerir una importante carga adicional para los administradores del repositorio, como por ejemplo al convertir archivos entre formatos o en el momento de verificar y corregir estos archivos. Al realizar estas actividades, es necesario minimizar tanto la dependencia con los administradores (o sea, qué nivel poseen tanto de conocimientos técnicos como del uso de las herramientas de transformación y verificación) así como también cualquier decisión sujeta a consideraciones subjetivas, como por ejemplo cuál es el nivel de calidad mínimo requerido para determinar que una transformación sea o no aceptable. De este modo, no solo se asegura un nivel de calidad mínimo general para todos los documentos, sino que también se facilita la incorporación o el intercambio de personas en el staff de administración del repositorio.

La política de preservación digital establecida en SEDICI requiere que la carga de objetos digitales al repositorio debe realizarse siempre en un formato preservable. Por otro lado y como es de esperarse, las actividades de catalogación y carga no se concentran en una única persona, sino que participan varias personas que realizan por lo general cualquiera de las tareas necesarias para obtener archivos aptos para su carga: división de documentos, transformación de formatos, corrección de metadatos de los documentos, incorporación de OCR, entre otras. Por este motivo, al diseñar el modelo de trabajo aquí descrito, se buscó implementar un sistema independiente de las personas y que promueva el trabajo colaborativo en equipo. Como ganancia adicional, se logró una implementación que puede accederse desde cualquier computadora de la red de trabajo, independiente de su sistema operativo, arquitectura o incluso capacidad de cálculo, y que evita la instalación de herramientas particulares en las computadoras de trabajo de los administradores, lo que agiliza la incorporación de equipos y brinda flexibilidad a la hora de cambiar equipos o realizar actualizaciones o cambios en los sistemas operativos subyacentes.

La propuesta de este trabajo se basa en un modelo de trabajo centralizado, que concentra el procesamiento de los documentos en un equipo dedicado, y desatendido, lo que significa que los administradores no necesitan intervenir en el procesamiento de dichos documentos, aunque se requiere su intervención a la hora de analizar y validar los documentos procesados, y eventualmente realizar operaciones adicionales para completar este procesamiento. La implementación de este modelo implica un sistema de red estilo cliente-servidor muy sencillo y fácil de replicar, en el que los clientes (los administradores) acceden a un directorio compartido en la red y depositan los documentos que deben ser procesados, que luego son tomados por un proceso (servidor) que los analiza, transforma y deposita en otro directorio de la red el resultado de esta transformación. Para la compartición de los directorios se utilizó la tecnología de directorios sobre redes TCP/IP de MS Windows conocida como Server Message Block (SMB, Microsoft 2016), que también está disponible en equipos con el sistema operativo MAC OS X y en sistemas



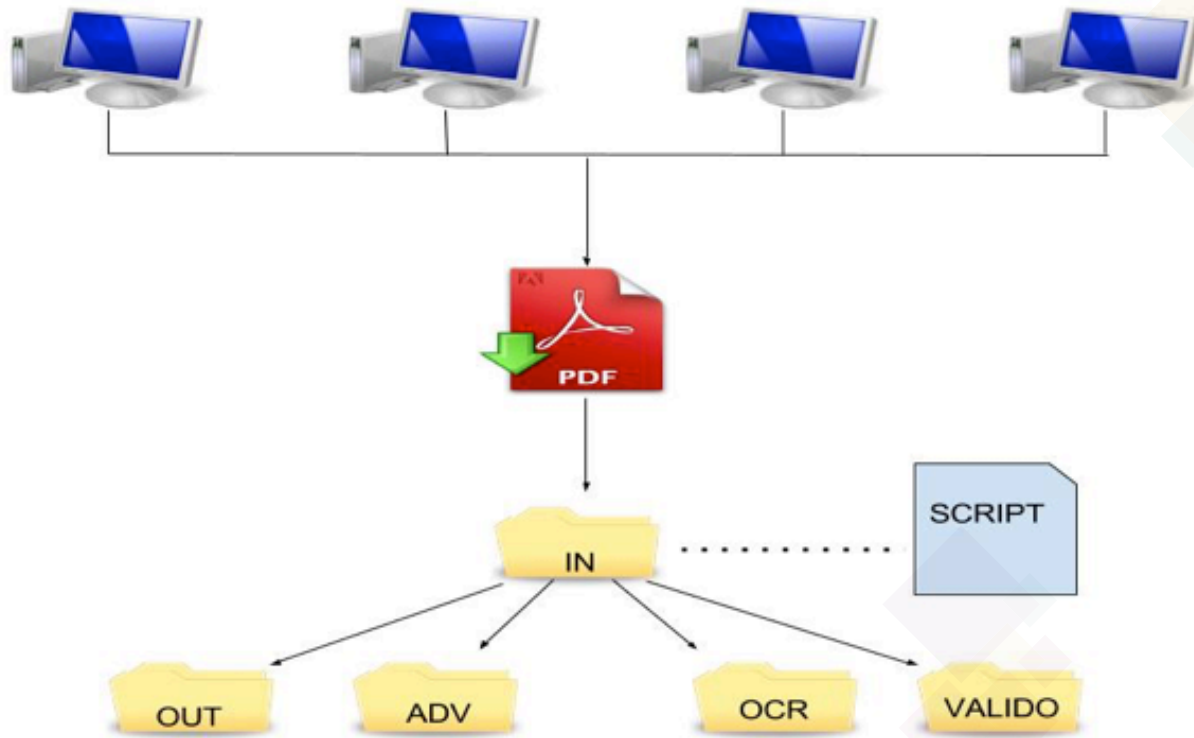
GNU/Linux mediante el cliente Samba (Samba 2016); de este modo, cualquier computadora de la red capaz de acceder a una carpeta compartida en la red podrá utilizar este servidor de transformaciones. Asimismo, dado que la propuesta es independiente del mecanismo de compartición de directorios en red, este mismo modelo puede también replicarse sobre el protocolo Common Internet File System (CIFS, 2016), lo que permite superar la red de área local y ampliar el espacio de trabajo hacia cualquier equipo conectado a Internet. Incluso es posible implementarse mediante otros sistemas de directorios compartidos, como Network File System (NFS, 2016) o utilizando aplicaciones en la nube como por ejemplo Owncloud (Owncloud, 2016) o Dropbox (Dropbox, 2016).

El espacio de trabajo (el directorio compartido) se organizó en distintas áreas o secciones, que internamente se corresponden con directorios del sistema de archivos. Esto permite organizar el trabajo ya realizado del trabajo que aún falta realizar, y ayuda a los administradores a comprender y utilizar este modelo. En una primera implementación, el modelo contaba con tres secciones:

- 1) Sección In:** espacio en donde los administradores depositan los archivos a convertir.
  - 2) Sección Out:** espacio donde son depositados los archivos convertidos satisfactoriamente, o sea que fueron transformados y verificados.
  - 3) Sección Advertencia:** espacio donde se colocan los archivos que no han pasado el validador post conversión.
- Si bien este esquema resultaba muy simple de utilizar, al revisar el modelo fue necesario incorporar dos secciones adicionales para manejar ciertos casos especiales:
- 4) Sección Válido:** directorio donde se copian los archivos que ya se encuentran en un formato de preservación (o sea, que no requieren ningún procesamiento) y que han pasado su correspondiente validación.
  - 5) Sección OCR:** directorio donde se depositan los documentos que directamente no se pueden convertir. Por lo general a estos documentos no se les ha realizado un reconocimiento de caracteres y posterior extracción de texto, requisito fundamental para los documentos de texto del repositorio.

Para los clientes o usuarios finales (los administradores que requieren las transformaciones), la metodología de trabajo es muy simple (imagen 1). Cuando se recibe un documento en formato PDF (por ejemplo una tesis, un artículo o un libro), el administrador a cargo del mismo accede al directorio compartido en la red, y copia el archivo original en la Sección In. Luego, espera a que el servidor realice las transformaciones necesarias, y a continuación ingresa a la Sección Out, donde encontrará tanto el documento transformado a un formato preservable junto a un archivo de registro de la transformación. En caso que el archivo obtenido no haya cumplido con la validación, se deberá buscar en la Sección Advertencia, y revisar el registro de transformación correspondiente para analizar qué ha sucedido y evaluar su reparación. Si el archivo está corrupto y no hay forma de convertirlo o repararlo, pasa a la sección OCR donde se realiza el reconocimiento óptico de caracteres. Este procesamiento se realiza al día de hoy con el software ABBYY FineReader (ABBYY, 2016), el cual una vez finalizado permite exportar el documento en formato PDF/A. Por otro lado, en caso que el archivo ya se encontraba en un estándar de PDF/A válido, el administrador lo encontrará en la Sección Válido, junto a un breve reporte donde se confirma la validez del mismo. En todos los casos, además del reporte

particular de transformación y validación que se obtiene por cada archivo, se registran los eventos en una bitácora general, que luego permite obtener estadísticas de uso de la aplicación (archivos procesados, transformaciones incorrectas, tipos de errores encontrados, etc). Este registro general sirve como retroalimentación para continuar mejorando y ampliando este modelo.



**Imagen 1: Esquema de trabajo general. Los administradores acceden a un directorio de red, donde depositan los documentos PDF, que luego son tomados por un script, procesados y almacenados en otro directorio según el resultado del procesamiento (OUT, ADV, OCR, VÁLIDO)**

Algunos OD no aprueban el post análisis porque no cumplen el estándar en el que se intentó convertir. Esto suele pasar cuando el algoritmo falla en la detección del nivel o de la conformidad si, por ejemplo, se intentó convertir un OD a PDF/A-1b, pero el archivo contiene imágenes con transparencias. En este ejemplo, un posible camino sería intentar convertir el OD a PDF/A-2b; otra posible opción sería eliminar las transparencias. Al día de hoy, este análisis y reparación se debe realizar de manera manual por los administradores, que encontrará estos “casos especiales” en la sección Advertencia. Sin embargo, en caso de que no haya un camino posible para una conversión exitosa, el archivo PDF deberá pasar a la sección OCR para procesarse con el software ABBYY.

Respecto a la conversión de PDF a PDF/A, se realizó un relevamiento sobre distintas herramientas en el que se consideraron diferentes parámetros como ser la licencia (software libre vs sistema cerrado), la interfaz de uso (línea de comandos, aplicación de escritorio y proceso servidor) y plataforma (MS Windows, GNU/Linux). La herramienta pdfaPilot resultó muy apropiada, pero su versión para servidores tenía un costo de adquisición mayor al presupuesto disponible, y por ellos se optó por adquirir la versión CLI

para GNU/Linux, cuyo uso se debe realizar por medio de un intérprete de comandos de dicho sistema operativo. Adicionalmente, por medio de un programa (script) escrito en lenguaje Bash, se encapsularon y se brindó acceso a los módulos encargados de realizar las distintas tareas sobre los documentos, como ser la detección de archivos pendientes de transformación, la conversión entre formatos, la validación, el registro de las actividades realizadas en la bitácora, la salida hacia el usuario, etcétera. Este script se desarrolló con un área dedicada a las variables de configuración del sistema (por ejemplo, niveles de registro o rutas hacia los programas ejecutables), lo que permite adaptar su funcionamiento en diferentes entornos de ejecución. El script en cuestión se ejecuta de manera continua en segundo plano, y cada vez que detecta que se han copiado archivos o directorios dentro de la Sección IN, ejecuta las aplicaciones que realizan los análisis y transformaciones correspondientes, y almacena los resultados en los directorios apropiados según el esquema previamente descrito.

El procesamiento recursivo de directorios es un arma de doble filo. Si bien puede ahorrar tiempo a los administradores, ya que podrían copiar un directorio con muchos subdirectorios dentro con archivos a procesar o con más subdirectorios, esto puede generar una sobrecarga de trabajo en el servidor, manteniendo procesos en ejecución por mucho tiempo y ocupando demasiados recursos de procesamiento. Por este motivo, en el módulo de procesamiento central de la herramienta se incorporó la posibilidad de copiar directorios enteros con archivos dentro de la sección IN, pero se limitó la recursión a un sólo nivel de acceso. Esto también acorta los tiempos de copiado de archivos en la red, lo cual podría generar otro punto de falla dado que los archivos sólo deben procesarse una vez que se han copiado completamente. Desde el punto de vista de la seguridad, se incorporaron dos módulos adicionales. El primero de ellos se encarga de brindar exclusión mutua sobre cada archivo, verificando que el mismo no esté siendo accedido por algún proceso de conversión previo. El segundo módulo está encargado de verificar que el archivo ingresado es un documento PDF correcto, o lo que es lo mismo, que no se ha corrompido al descargarse o copiarse a través de la red.

El programa pdfaPilot permite que se le especifique el nivel (1, 2 o 3) y el nivel de conformidad buscado (a,b,u). Para hacer uso de estas opciones, se encapsuló en un módulo la configuración de la ejecución de las transformaciones, en el cual se determina cuál es estándar más indicado para cada PDF. Para ello, previamente se realiza un análisis de los archivos PDF, mediante otras herramientas como exiftool o jhove que, entre otras opciones, permiten extraer los metadatos contenidos dentro de los archivos PDF. Un simple análisis de los metadatos extraídos permite determinar el nivel de conformidad alcanzable; por ejemplo, si el documento ingresado no es “taggeado” (etiquetado), no será posible alcanzar la conformidad “a” de PDF/A, y por lo tanto se optará por la conformidad “b”.

Todas las decisiones tomadas, los resultados alcanzados, el motivo y los errores encontrados durante el procesamiento deben ser informados de alguna manera a las personas que solicitan el procesamiento de los archivos. Por este motivo, se incorporó un módulo encargado de registrar los eventos sucedidos y generar un reporte final para los usuarios. Este módulo, además de incorporar el reporte generado por el software convertidor, genera tres tipos de reportes:

1. Reporte estadístico general: Archivo de texto donde, para cada OD detectado en la Sección In, se agrega el nombre del archivo, fecha y resultado del procesamiento, es decir, si el archivo terminó en la Sección Out, Válido, Ocr o Advertencia
2. Reporte del documento: por cada archivo procesado se crea reporte con una breve descripción de lo que la herramienta realizó, su resultado post-conversión y un código informativo (imágenes 2, 3 y 4).
3. Reporte diario: Archivo de texto generado por día con una síntesis de los OD procesados.

```

04_Alimentación y n...ión.pdf-PDFA-log.txt x
- Opening file 04_Alimentación y nutrición.pdf.
- Analyzing 04_Alimentación y nutrición.pdf.
- Font 'GeomaLightDemo' not found and substituted with 'multiple master font'.
- Font 'GeomaRegularDemo' not found and substituted with 'multiple master font'.
- Font 'MindBlue' not found and substituted with 'multiple master font'.
- Conversion events:
  - Font substituted.
- Performing post analysis for 04_Alimentación y nutrición.pdf-PDFA.pdf.
* Post analysis errors in 04_Alimentación y nutrición.pdf-PDFA.pdf.

Codigo: 6

```

*Imagen 2: La imagen muestra el reporte de conversión del archivo "04\_Alimentación y nutrición.pdf". Falla el post-análisis por lo que el archivo original junto a su reporte son enviados a la Sección Advertencia (código de error 6).*

```

report-File-In.txt x
- Opening file In.pdf.
- Analyzing In.pdf.
- Copied output intent from input file.
- Performing post analysis for In-PDFA.pdf.
- Post analysis for In-PDFA.pdf has been successful.
- File In.pdf converted successfully.

Codigo: 5

```

*Imagen 3: Reporte de conversión de un archivo satisfactorio (código de aceptación 5). El OD original, el OD convertido a PDF/A y el reporte son enviados hacia la Sección Out.*

<pre> Valido.txt x - Opening file Valido-PDFA.pdf. - Analyzing Valido-PDFA.pdf. - File Valid-PDFA.pdf converted successfully.  Codigo: 0 </pre>	<pre> out-log.txt x - Opening file error.pdf. * Cannot open file error.pdf.  Codigo: 1 </pre>
---	---

*Imagen 4: El reporte de la izquierda corresponde a un documento que ya se encontraba en formato PDF/A válido, por lo que es llevado a la Sección Valido, con código de aceptación 0. El reporte de*



la derecha corresponde a un archivo que no se pudo procesar por un algún error, generalmente, el archivo está malformado o corrupto, por lo que es enviado a la Sección OCR con código de error 1.

## Mejoras y actualizaciones

Durante la primera etapa de uso de este desarrollo, se observó que la detección del nivel y conformidad de PDF/A no resultaba apropiada en todos los casos (estaba muy sujeta a las características internas de cada archivo a procesar). Sumado a esto, el tiempo de transformación de estos archivos se incrementó de manera considerable, pues se repetían numerosas veces los procesos de análisis y conversión, hasta obtener finalmente un resultado fallido. Por estos motivos se buscó una segunda herramienta capaz de realizar estas tareas, o que permitiera dar un nivel y conformidad por defecto, y que sea capaz de alterar el nivel o la conformidad buscados en caso de ser necesario. Se optó por la herramienta 3Height converter shell (3H, imagen 5) para GNU/Linux, que además de contar con estas opciones, la empresa desarrolladora (PDF-tools) proveyó un buen servicio de servicio técnico y otorgó un importante descuento de adquisición por tratarse de una institución educacional. El cambio de aplicación no requirió mayores modificaciones en el script de ejecución, ya que sólo fue necesario actualizar la ruta del programa ejecutable y modificar los parámetros según son requeridos por la nueva herramienta.

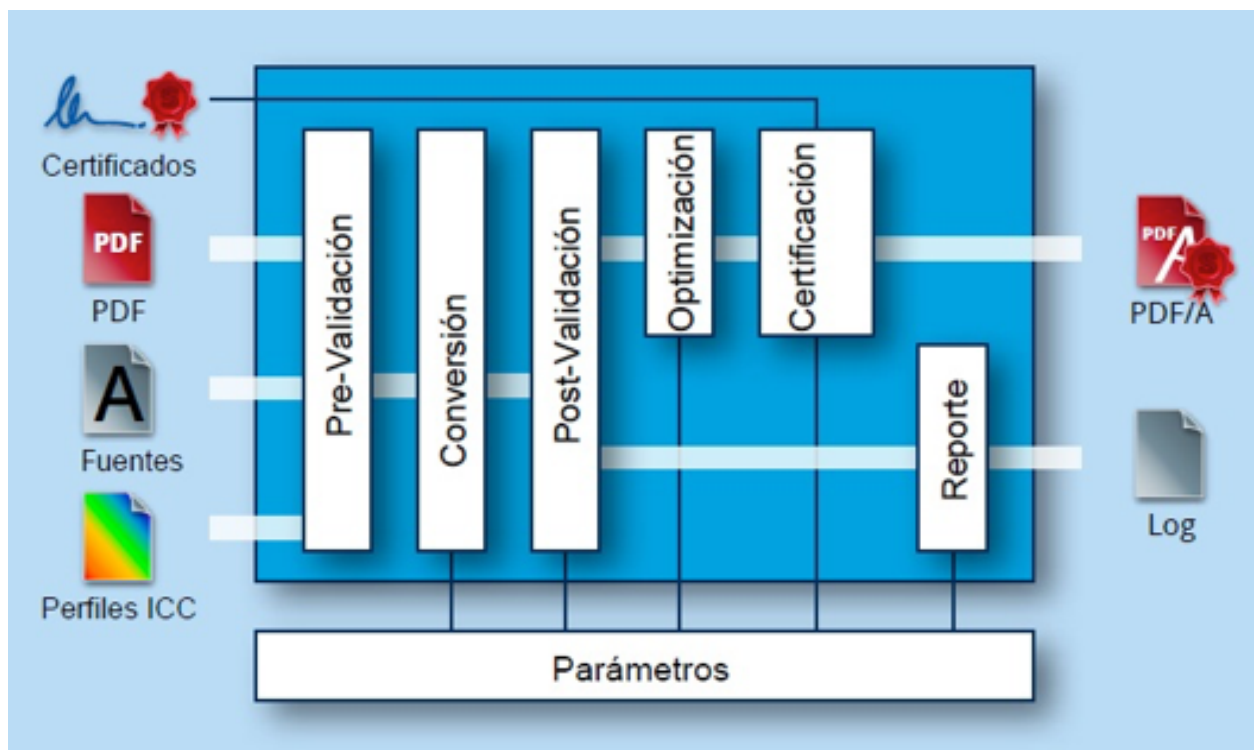


Imagen 5: Esquema de procesamiento por etapas de 3H (Fuente: manual de usuario de 3H, traducido al castellano).

Una de las especificaciones del estándar PDF/A establece que las fuentes deben incluirse (embeberse) en los documentos; dicho de otro modo, no se aceptan fuentes externas. La elección de la fuente a utilizar presenta algunos inconvenientes, principalmente cuando se trata de fuentes privativas o fuentes no disponibles en todas las plataformas y sistemas operativos. Para resolver esta situación, se sumó un conjunto de paquetes de fuentes (por ejemplo fuentes TrueType) al conjunto de fuentes por defecto, con lo cual en caso de no encontrarse una fuente específica, podrá optarse por una versión alternativa

del paquete adicionado. Esto requirió unas modificaciones en el software 3H por parte de la empresa desarrolladora, lo que generó una nueva versión de esta aplicación. Otras situaciones también requirieron intercambios con el soporte técnico. Como se mencionó previamente, si no se especificaba un nivel y conformidad, el software seleccionaba aquel que parecía correcto mediante su algoritmo de decisión, pero en algunos casos el OD convertido era nivel “2” conformidad “u”, cuando el mismo cumplía los requerimientos para ser “1a”. Al observar este comportamiento, se notificó al soporte técnico, quienes generaron una nueva versión incorporando un provisorio “upgrade detection” más exacto. En la versión actual estos asuntos fueron resueltos mediante la incorporación de dos parámetros adicionales, uno que sube el nivel (de 1 a 2) y otro que baja la conformidad (de “a” a “b” en caso de nivel 1, y de “a”, “u”, “b” en nivel 2).

Desde el punto de vista de la validación de los documentos convertidos, se realizaron pruebas tanto con 3H (en su última versión) como con el validador de Acrobat Pro DC. Si bien en líneas generales ambos funcionaban correctamente, se encontraron dos metadatos que eran tratados de manera diferente. Por un lado, la forma de definir los metadatos de fecha no era detectada apropiadamente por Acrobat DC. Se notificó al servicio técnico de Adobe con respecto al XMP metadata Create Date, que debe especificarse en el estándar ASN.1 como se indica en la norma ISO/IEC 8824; el soporte técnico confirmó que el mismo será corregido en una futura versión del software. Finalmente, se encuentra abierto el debate acerca del contenido del metadato History, ya que es interpretado de manera diferente por 3H y por Acrobat DC, y el estándar no aporta claridad al respecto.

## **Conclusiones**

Una de las ventajas del modelo aquí presentado yace en la flexibilidad que ofrece en cuanto al trabajo de los administradores y la independencia de sus estaciones de trabajo respecto a las herramientas de procesamiento de documentos. En contextos donde surgen constantemente nuevas plataformas y formas de acceso y compartición de recursos, este tipo de flexibilidad sirve de apoyo para continuar expandiendo los equipos de trabajo o adaptando los entornos según las necesidades de cada usuario.

Por otro lado, el modelo propuesto se destaca por su facilidad de replicación en otras organizaciones. Las tecnologías que utiliza son compatibles con la mayoría de las plataformas, y en las estaciones de trabajo sólo es necesario acceder a un directorio compartido en red. Desde el lado del servidor, el script desarrollado ofrece un conjunto de parámetros de configuración, lo que demostró ser muy útil al cambiar de herramientas y por lo tanto los parámetros de ejecución de las mismas (de pdfaPilot a 3H), y que servirá también para la instalación en otros contextos. En lo relativo a los costos, si bien se adquirió una licencia de software para el software 3H, se obtuvo un importante descuento que permitió mantener el proyecto dentro del presupuesto. Asimismo, muchas de las herramientas utilizadas son de código abierto (servidor Linux, Samba, exiftool, jhove), lo que también asegura que los costos se mantendrán al mínimo a lo largo del tiempo. Particularmente en este trabajo con respecto al software de 3H, solamente se mencionó de su módulo de conversión y validación, pero el mismo cuenta con más opciones de configuración. Por ejemplo, una de ellas le aporta al PDF/A creado la optimización (linealizado o con vista rápida para la web). Con Vista rápida en Web, el servidor Web sólo carga la página solicitada, en lugar del PDF completo. Esta opción es deseada en documentos gran tamaño, que pueden tardar mucho tiempo en descargarse desde un servidor.

En cuanto a los tiempos de procesamiento, la conversión y validación de un documento PDF tradicional es casi instantáneo (aproximadamente 1 segundo). La revisión de la Sección IN para detectar OD a procesar se realiza cada 10 segundos (este valor puede configurarse también desde el script). Con estos valores temporales, en la mayoría de los casos los administradores copian archivos y directorios en la sección IN, y obtienen los documentos transformados en menos de 20 segundos, lo cual agiliza enormemente la tarea de carga. Si bien se requieren conocimientos más avanzados en los distintos estándares de PDF/A para la reparación de “casos especiales”, en la mayoría de los casos esto no sucede y, por lo tanto no se ha generado un cuello de botella por el tratamiento de estos documentos. Esta necesidad de acción manual en ciertos casos presenta uno de los principales puntos a mejorar, pues sería ideal lograr un procesamiento totalmente automático o al menos minimizar tanto como sea posible los casos particulares que requieren intervención humana. Esta automatización puede alcanzarse mejorando los algoritmos de elección de estándares o incorporando nuevas herramientas de detección y reparación, y será uno de los trabajos a futuro del proyecto.

La incorporación de fuentes presenta también algunos desafíos particulares. Día a día se crean nuevas fuentes, algunas de ellas abiertas pero otras privativas, y en muchos casos no disponibles en todas las plataformas. Se hace muy difícil contar siempre con todas las fuentes requeridas, por lo que en ocasiones se opta por una fuente alternativa, que puede no ser igual a la fuente original del documento. Si bien esto asegura que el contenido del documento (el texto por ejemplo) será accesible a lo largo del tiempo, que es el fin último de estos esfuerzos por asegurar la preservación digital, sería deseable mantener los documentos transformados tan parecidos a los originales como sea posible. Quizás este punto requiera un trabajo de concientización de los usuarios, para que minimicen o eviten el uso de fuentes incompatibles o privativas, pero esto ya excede los objetivos del proyecto.

Es posible también considerar la escalabilidad de esta propuesta. Actualmente el procesamiento de los OD se realiza de manera individual y secuencial (uno detrás del otro). Sería deseable paralelizar estas tareas, ya sea para aprovechar las ventajas de los equipos modernos con múltiples unidades de procesamiento, o incluso para distribuir la carga de trabajo en equipos de la red. Si bien al día de hoy esto no es un problema en el contexto de SEDICI, el crecimiento sostenido del repositorio o incluso la realidad de otras organizaciones y contextos amerita considerar seriamente una solución que permita multiplicar la carga de trabajo por varias veces sin que esto repercuta en la eficiencia del servicio. Una observación no menor respecto a esta propuesta es que no se encuentra integrada con el software Dspace que utiliza el repositorio. Los administradores deben descargar los documentos desde Dspace, procesarlos por fuera, y luego cargar las versiones ya procesadas. Sería ideal que esto se realice de manera transparente, y que los administradores dispongan directamente de los documentos originales y procesados, y se centren en la descripción de los mismos. Esto requerirá probablemente modificaciones en el software Dspace, así como también evaluar nuevamente el flujo de trabajo de los administradores del repositorio, pero sin dudas ofrecerá grandes ventajas para el trabajo diario del repositorio.

Para finalizar, cabe destacar que el proyecto completo se encuentra disponible en el repositorio Github, para que cualquier organización pueda descargar los scripts, utilizarlos o personalizarlos según sus necesidades particulares. La URL del proyecto en Github es <https://github.com/sedici/scripts-digitalizacion>.

## Bibliografía

Brown, A. (2008) Digital Preservation Guidance Note 1: Selecting file formats for long-term preservation. The National Archives. Recuperado en <https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>

CCSDS, Reference Model for an Open Archival Information System (OAIS):ISO 14721 . 2002.

Common Internet file system <https://technet.microsoft.com/en-us/library/cc939973.aspx>

De Giusti M. R., Sobrado A., Lira A. J., Vila M. M., and Villarreal G. L. (Sep. 2008) "SeDiCI (Servicio de Difusión de la Creación Intelectual)," D-Lib Magazine, vol(14).

De Giusti M.R, Lira A. J., Texier J., and Villarreal G. L. (2012) "Las actividades y el planeamiento de la preservación en un repositorio institucional". Conferencia Internacional BIREDIAL-ISTEC, Universidad del Norte, Barranquilla - COLOMBIA. Recuperado en <http://sedici.unlp.edu.ar/handle/10915/26045>

De Giusti, M. R., Lira, A. J., Villarreal, G. L., Terruzzi, F. A.; Adorno, F. G. (2014) Preservación digital: un experimento con SEDICI-DSpace. XX Asamblea General de ISTEC (Puebla, México). Recuperado en <http://sedici.unlp.edu.ar/handle/10915/34889>

De Giusti, M. R.(2014) "Una metodología de evaluación de repositorios digitales para asegurar la preservación en el tiempo y el acceso a los contenidos." Tesis doctoral, Universidad Nacional de La Plata. Recuperado en <http://sedici.unlp.edu.ar/handle/10915/43157>

Descarga la versión de prueba de adobe acrobat | acrobat pro DC <https://acrobat.adobe.com/la/es/free-trial-download.html>

Dropbox (IE) <https://www.dropbox.com/>

ExifTool by Phil Harvey <http://www.sno.phy.queensu.ca/~phil/exiftool/>

Giménez Chornet, V. (2014). Criterios ISO para la preservación digital de los documentos de archivo. Códices, 10(2), 135-150. Recuperado de <http://revistas.lasalle.edu.co/index.php/co/article/view/3267/2607>

ISO 19005-1. Document management - Electronic document file format for long-term preservation - Part 1: Use of PDF 1.4 (PDF/A)

ISO 19005-2. Document management - Electronic document file format for long-term preservation - Part 2: Use of ISO 32000-1 (PDF 1.7)

ISO 19005-3. Document management - Electronic document file format for long-term preservation - Part 3: Use of ISO 32000-1 with support for embedded files (PDF/A-3) PDF/A-3: ISO 19005-3

ISO 32000-1:2008. Document management -- Portable document format -- Part 1: PDF 1.7

JSTOR/Harvard object validation environment <http://jhove.sourceforge.net/>

Kresse W., Pau J.M., "Development of an ISO-standard for the preservation of geospatial data and metadata: ISO 19165," Photogrammetrie, Fernerkundung, Geoinformation, vol. (Dec. 2015), Pages 449-456.

Linux NFS faq <http://nfs.sourceforge.net/>

Microsoft SMB protocol and CIFS protocol overview [https://msdn.microsoft.com/en-us/library/windows/desktop/aa365233\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/aa365233(v=vs.85).aspx)

Microsoft typography - OpenType specification <https://www.microsoft.com/en-us/Typography/OpenTypeSpecification.aspx>

OCR software from ABBYY. Best text recognition for windows and Mac <https://www.abbyy.com/finereader/>

Opening windows to a wider world <https://www.samba.org/>

OwnCloud.org <https://owncloud.org/>

PDF to PDF/A converter <http://www.pdf-tools.com/pdf/pdf-to-pdf-a-converter-signature.aspx>

PREMIS, "PREMIS: Preservation Metadata Maintenance Activity (Library of Congress)," (2016). Recuperado en <http://www.loc.gov/standards/premis/>

Products | Callas Software. <https://www.callassoftware.com/en/products/pdfapilot>



# Interoperabilidad

Meilyn Garro Acón (UCR)

# “Propuesta de un esquema de metadatos, a partir del crecimiento del estándar de Dublin Core, para el Repositorio Universitario Especializado en Comunicación Pública de la Ciencia”

*Autores\*: Carmina de la Luz Ramírez<sup>1</sup>, Raúl Ortega Muñoz<sup>2</sup>, Óscar Hernández Hernández<sup>3</sup>, Maribel González González<sup>4</sup>, Addina Cuervo Espinosa<sup>5</sup>, Esteban López Jiménez<sup>6</sup> y Roberto Iván Ramírez García<sup>7</sup>*

*\*Dirección General de Divulgación de la Ciencia, Universidad Nacional Autónoma de México (DGDC, UNAM). Circuito cultural de Ciudad Universitaria, Coyoacán, C. P. 04510, Cd. Mx., México.*

---

## **Resumen:**

Los repositorios digitales son sitios web donde se centraliza, almacena y mantiene información digital, y hoy en día constituyen lugares virtuales que resguardan la producción intelectual de las instituciones al tiempo que brindan el acceso libre y gratuito a los materiales por parte de cualquier persona con conexión a Internet. Actualmente, la Universidad Nacional Autónoma de México cuenta con una red de repositorios que requiere esfuerzos en términos de visibilidad, interoperabilidad y aprovechamiento. Ameyalli-Repositorio Universitario Especializado en Comunicación Pública de la Ciencia es uno de estos repositorios. El presente trabajo tiene el objetivo de compartir el proceso de desarrollo de un esquema de metadatos, basado en Dublin Core e implementado en DSpace, para la creación de Ameyalli. La base de esta propuesta es una metodología híbrida de administración de proyectos. El esquema de metadatos fue desarrollado en diversas etapas, en las cuales jugaron un papel trascendental los objetivos y lineamientos de repositorio, la producción intelectual de la institución y los usuarios finales de los recursos digitales. El principal resultado es un esquema compuesto por 40 elementos y subelementos en Dublin Core. Este se ha convertido en una referencia para la normalización de metadatos dentro de la UNAM, dirigida a mejorar la interoperabilidad de los acervos digitales de dicha universidad.

**Palabras clave:** Ameyalli, repositorio digital, comunicación pública de la ciencia (CPC), DSpace, metadatos, Dublin Core (DC)

## **Abstract:**

Digital repositories are websites where digital information can be centralized, stored and maintained. Today, institutions protect their intellectual production through digital repositories and, at the same time, provide for anyone who has Internet connection free access

to the resources. The National Autonomous University of Mexico (UNAM, for its acronym in Spanish) has a network of digital repositories, which requires efforts in terms of visibility, interoperability and use. Ameyalli-Specialized in Public Communication of Science University Repository is one of these repositories. This paper shares the process of developing a metadata schema based on Dublin Core and implemented in DSpace for Ameyalli. The basis of this proposal is a hybrid project management methodology. The metadata schema was developed in several stages, where Ameyalli's objectives and guidelines, intellectual output of the institution and the users of the repository played an important role. The main result of this work was a 40 elements and subelements scheme in Dublin Core, which has become into a reference in order to establish a global scheme within University and to improve UNAM's repositories interoperability.

**Key words: Ameyalli, digital repository, public communication of science (PCS), DSpace, metadata, Dublin Core**

## Introducción

Los repositorios digitales son sitios web donde se centraliza, almacena y mantiene información digital, como bases de datos y diversos archivos (Universidad de Salamanca, 2016). Sus orígenes se remontan a la necesidad por parte de las instituciones, principalmente universitarias, de compartir el conocimiento generado por y entre sus académicos, desencadenando el desarrollo de soluciones de software que facilitan el acceso y administración a una gran cantidad de información en distintos formatos, de tal forma que actualmente los repositorios digitales constituyen lugares virtuales donde es posible depositar y resguardar la producción intelectual de las instituciones al tiempo que brindan el acceso libre y gratuito a los materiales por parte de cualquier persona con conexión a Internet (DGTIC, 2010).

Por otro lado, la Universidad Nacional Autónoma de México (UNAM) es considerada uno de los centros de educación superior más grandes del mundo (QS Universities Ranking, 2016), y debido a su amplio espectro de actividades, su numerosa comunidad y abundante producción de contenidos, se está enfrentando a una demanda excesiva de acceso a estos (López y García, 2007). Para resolverlo, la UNAM echó a andar el proyecto de investigación piloto, estudiando las tecnologías y prácticas necesarias para el desarrollo de una red de repositorios; esto en el marco del Macroproyecto para las Tecnologías de la Información y la Comunicación 2005-2009 (Galina y Jiménez, 2010). A su vez, dicho macroproyecto se inserta en una política de la UNAM dirigida a brindar a la sociedad plenas condiciones de acceso al conocimiento, acervos culturales y diversos servicios (UNAM, 2014).

Actualmente, la UNAM cuenta con numerosos acervos digitales, pero de acuerdo con Galina y Jiménez (2010) aún se requieren mayores esfuerzos e iniciativas para mejorar su visibilidad, representatividad y aprovechamiento, así como su interoperabilidad. Dentro de este contexto surgió Ameyalli-Repositorio Universitario Especializado en Comunicación Pública de la Ciencia, proyecto sustantivo de la Dirección General de Divulgación de la Ciencia (DGDC) de la UNAM. Los principales retos que implicó la creación de Ameyalli, sobre todo en términos de interoperabilidad fueron:

1. Selección del software apropiado (DSpace)
2. Implementación de un esquema de metadatos (Dublin Core)

Ninguno de estos dos asuntos es trivial, pues sus respectivas soluciones tuvieron que llevarse a cabo en función de su misión, visión, objetivos y lineamientos. Por ello, el presente trabajo está dirigido a compartir el proceso de desarrollo de un esquema de metadatos, basado en Dublin Core e implementado en DSpace, para la creación de Ameyalli-Repositorio Universitario Especializado en Comunicación de la Ciencia.

Dublin Core es un modelo de metadatos cuyo objetivo es representar de manera estandarizada cualquier información acerca de recursos digitales, a través de 15 definiciones semánticas llamadas elementos (ver estos elementos estándar en el cuadro de la página 8), los cuales se dividen en tres grupos que indican la clase o el ámbito de la información que se guarda en ellos: elementos relacionados con el contenido del recurso, elementos relacionados con la propiedad intelectual del recurso, y elementos relacionados con su instanciación. Estos pueden ser calificados, creando subelementos que proporcionan información más detallada del recurso.

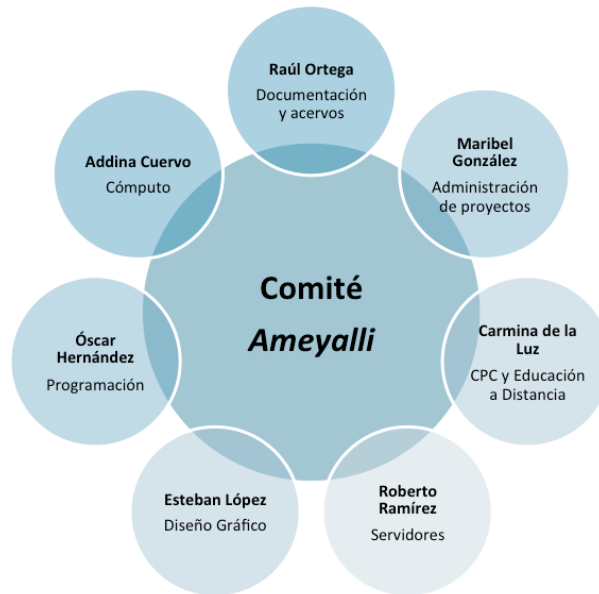
La Iniciativa Dublin Core (DCMI) comenzó en 1995. Actualmente está disponible en 20 idiomas, ha sido adoptado por el CEN/ISS (European Committee for Standardization / Information Society Standardization System) y es también estándar oficial del WWW Consortium (Valencia-Velázquez, 2013). Dublin Core es manejado de forma predeterminada por DSpace, lo que permite crear comunidades, colecciones e ítems que constituyen una taxonomía adecuada a los objetivos del repositorio y necesidades de los usuarios. Además, Dublin Core es interoperable con otros repositorios, tanto en DSpace como en otros tipos de software.

## **Materiales y metodología**

Con el fin de asegurar lo mejor posible el éxito del proyecto, el primer paso en la creación de Ameyalli fue establecer una metodología “híbrida” de administración de proyectos, la cual combina los fundamentos propuestos por el Instituto de Administración de Proyectos y controles institucionales de la DGDC (Mendoza, 2007; PMI, 2016).

La propuesta metodológica “híbrida” se considera necesaria como instrumento de gestión y control, tomando en cuenta las fortalezas y debilidades en el éxito de nuestro proyecto para la creación del Repositorio Universitario Especializado en Comunicación Pública de la Ciencia-Ameyalli, para ello fue necesario analizar diferentes aspectos, pero los dos más importantes fueron, el valor que nos brindaba una metodología reconocida a nivel mundial del Project Management Institute, a través de su guía Project Management Book Of Knowledge (PMBOK y PMI, respectivamente), y como segundo el contexto en el que sería implementada de acuerdo con la diversidad de formas de trabajo del grupo multidisciplinario formado por el Comité de Ameyalli y en la propia dependencia. Otros aspectos considerados fueron los usuarios potenciales a los que está dirigido el proyecto, los recursos financieros, lineamientos y procesos institucionales ya establecidos tanto en la DGDC como en la UNAM.

En dicho Comité se estructuraron los roles y responsabilidades con respecto a los perfiles y áreas de conocimiento de los integrantes para gestionar el proyecto de Ameyalli, considerando las necesidades particulares del proyecto con el fin consolidar compromisos, responsabilidades, contribuciones, involucramiento y respaldo dentro del equipo de trabajo multidisciplinario (ver siguiente figura).

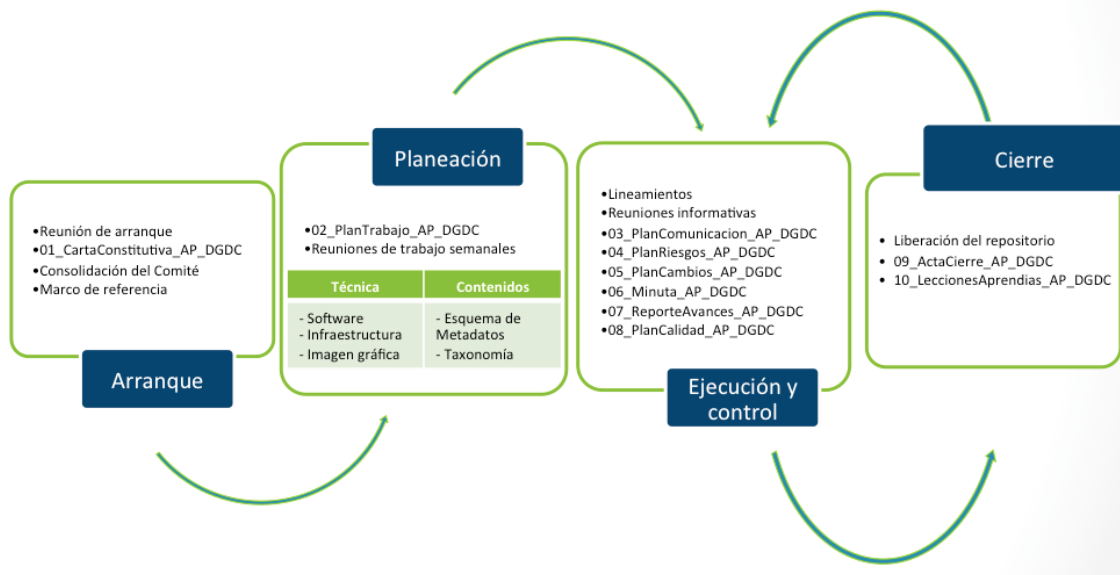


Y con el fin de no imponer un cambio radical y rígido en las formas trabajo y estandarizar el mecanismo de control y seguimiento bajo algo ya probado como el PMI (sin abordar todas sus etapas y gestiones de conocimientos), y la experiencia de cada miembro del Comité de Ameyalli, se acordó establecer esta propuesta metodológica “híbrida” para el seguimiento integral del proyecto, logrando fusionar procesos y plantillas del PMI con las buenas prácticas de gestión, respetando los lineamientos establecidos y marcados en el ámbito de la DGDC con un enfoque ágil, tradicional e integral.

Esta metodología también nos permitió la alineación del proyecto no sólo en proporcionar, estandarizar y centralizar documentación, procesos e información requerida y necesaria del proyecto, sino también mantener una línea directa en la comunicación entre los involucrados, durante todo su ciclo de vida (ver siguiente figura), alertando posibles riesgos y cambios relevantes, estableciendo acuerdos, compromisos y avances.

Dentro del seguimiento de esta implementación metodológica “híbrida”, se realizaron etapas y actividades sistemáticas de manera tal que los cambios, riesgos y mejoras se registraron en el proyecto, y al final de éste permitió tener un contexto general para documentar las lecciones aprendidas.

## Ciclo de vida de *Ameyalli*

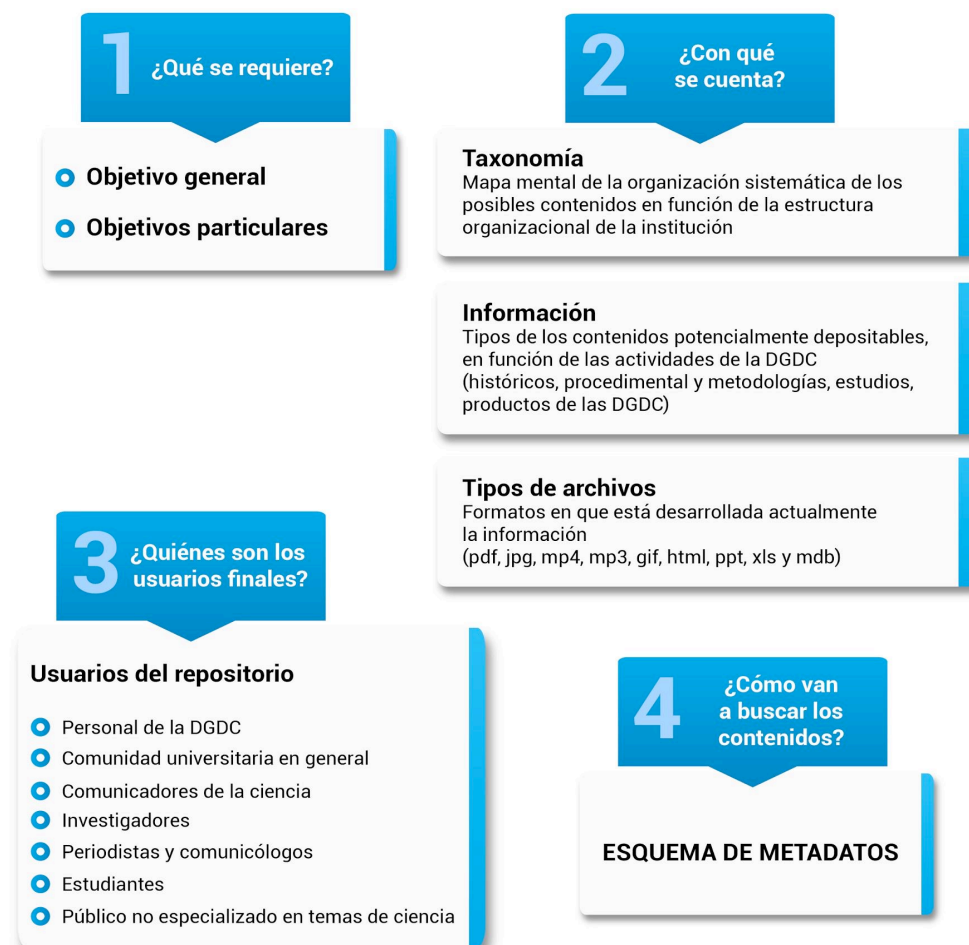


Esta propuesta de metodología “híbrida” hace hincapié en señalar que en los proyectos pueden existir fracasos por diversas razones: adaptarse a nuevas formas de trabajo, detalles técnicos, de presupuesto, de tiempos, desvió de objetivos, roles y funciones no claras, problemas de comunicación, etc.

Por otro lado, la selección de DSpace como software para la integración de los distintos elementos del repositorio se llevó a cabo debido a que este software resulta ser la herramienta idónea para el proyecto, no solo por su popularidad dentro de las instituciones académicas, sino porque es capaz de satisfacer las necesidades de los usuarios finales de Ameyalli (Dura Space, 2016). La selección del software se basó en buena parte en la experiencia que ya se tenía de un proyecto anterior, y sobre todo porque DSpace se sigue actualizando y ofreciendo nuevas herramientas para explotar el potencial de un repositorio, además de que permite la interoperabilidad, habilidad muy importante para los sistemas de información.

Por último, dado que DSpace utiliza por default el formato de metadatos de Dublin Core, se procedió a hacer un crecimiento de los 15 elementos estándar, tal y como se muestra en la siguiente figura (CDMI, 2016).





## Resultados

A grandes rasgos, el resultado preliminar del crecimiento del estándar de Dublin Core para Ameyalli, repositorio institucional de la DGDC, instalado en DSpace, es una propuesta de esquema de metadatos con más de 40 elementos y subelementos, pensado para que el público meta del repositorio, es decir, los distintos tipos de usuarios, encuentre los contenidos que busca (ver cuadro en la página 9).

Por otro lado, el esquema responde a la diversidad de actividades y productos de la DGDC que serán compartidos a través del repositorio, los cuales van desde aquellos relacionados con la museología y museografía, medios de comunicación (video, televisión, radio, prensa, redes sociales, portales de Internet), periodismo científico y literatura, for-

mación e investigación en comunicación pública de la ciencia, entre otros.

Otro resultado importante es que esta propuesta de esquema de metadatos está sirviendo como referencia para la normalización de todos los repositorios de la UNAM, actuales y futuros. Al respecto, se está trabajando dentro de una comisión de la Red de Acervos Digitales (RAD) de la máxima casa de estudios, en la cual también participa la Dirección General de Cómputo y de Tecnologías de la Información y Comunicación (DGTIC, UNAM). La normalización es de suma importancia porque permite comprender mejor cada definición guardada en los metadatos de los distintos ítems de cada repositorio, además de facilitar la interoperabilidad entre estos últimos, dando como resultado una red de repositorios con una mejor estructura en términos de sus metadatos, y que al finalizar contará con la documentación de sus procesos.

Contenido	Propiedad intelectual	Instanciación	Institucional
<b>Title</b> (título) - alternative (título alternativo) <b>Subject</b> (palabras clave) <b>Description</b> (descripción) - provenance (procedencia) - abstract (resumen) - table of contents (tabla de contenidos) <b>Source</b> (fuente) <b>Language</b> (idioma) <b>Relation</b> (relación) <b>Coverage</b> (cobertura) - espacial (espacial) - temporal (temporal)	<b>Creator</b> (autor) <b>Publisher</b> (editorial) <b>Contributor</b> (colaborador) - advisor (asesor científico) - broadcaster (locutor) - director (director) - editor (editor) - illustrator (ilustrador) -	<b>Date</b> - created (fecha de creación) - issued (primer fecha de publicación) - accessioned (fecha de registro en el repositorio) - available (fecha de disponibilidad en el repositorio) <b>Type</b> (tipo de documento) <b>Format</b> (formato del archivo) - extent (extensión, temporal o espacial) - médium (medio físico del documento) <b>Identifier</b> - URI - URL - ISBN - ISSN - DOI	<ul style="list-style-type: none"> <li>· Identificador interno (dgdcschema.id)</li> <li>· Clasificación según la oferta educativa de la UNAM (dgdcschema.classification)</li> <li>· Área que produce el contenido (dgdcschema.department)</li> <li>· Depositario del contenido en el repositorio (dgdcschema.depositary)</li> <li>· Depositor del contenido (dgdcschema.depositor)</li> <li>· Tipo de información del contenido (dgdcschema.information)</li> </ul>

En este cuadro, se marcan con negritas los 15 elementos estándar de Dublin Core, mientras que los cualificaciones que se derivan del crecimiento y adecuación del esquema estándar aparecen en azul oscuro. Por otro lado, en la última columna se agregaron los elementos de metadatos creados ex profeso para el repositorio institucional de la DGDC, esenciales para el control y documentación interna.

## Conclusiones

1. Los repositorios digitales inciden en las tres tareas sustantivas de las universidades públicas mexicanas: investigación, docencia y difusión de la cultura.

2. En el caso particular del repositorio Ameyalli, se considera un proyecto de éxito ya que el material que contendrá se ha desarrollado en la DGDC por más de 22 años, pueda ser consultado y revisado en diversas plataformas (móviles, tabletas, páginas, computadoras de escritorio, equipo portátil, etc).
3. El éxito de cualquier proyecto depende de la integración de una metodología para su administración y de la consolidación de un equipo de trabajo, mucho del éxito de este repositorio, se debe en gran medida a las diversas opiniones, conocimientos y experiencias de los integrantes del comité que lo forma.
4. El llevar una metodología “híbrida” de administración de proyectos, permitió llegar a los objetivos esperados en tiempo y forma.
5. El consolidar el formato de metadatos como Dublin Core permitió mejorar la visibilidad, representatividad y aprovechar la interoperabilidad con otros repositorios digitales.
6. El esquema de metadatos de un repositorio institucional debe construirse en función de sus lineamientos, objetivos, misión y visión.
7. El desarrollar un plan de trabajo con la metodología “híbrida” permitió que la RAD delegara y se apoyara en los integrantes del comité para iniciar la estandarización de metadatos a nivel institucional, y que otros repositorios lo tomen como referencia.

## **Bibliografía**

Dirección General de Cómputo y de Tecnologías de Información y Comunicación (DGTIC), UNAM (2010). Los repositorios digitales dentro de la UNAM. Consultado 25-junio-2016 en: <http://studylib.es/doc/316102/repositorios-digitales-en-la-unam>

# “PRÁTICAS DE INTEROPERABILIDADE PARA A INTEGRAÇÃO DE ACERVOS DIGITAIS HÉTEROGÊNEOS E DISTRIBUÍDOS”

•Autores: *Caterina Groposo Pavão; Janise Silva Borges da Costa; Manuela Klanovicz Ferreira; Zaida Horowitz*

---

## RESUMO

O trabalho aborda a interoperabilidade do Lume, Repositório Digital da Universidade Federal do Rio Grande do Sul (UFRGS), com outros sistemas de informação da Universidade, com os objetivos de automatizar e impulsionar o seu povoamento com documentos digitais produzidos na Universidade, minimizar inconsistências, evitar o retrabalho e, conseqüentemente, manter a integridade e uniformidade dos dados. Descreve as implementações realizadas no Lume, para permitir a sincronização com fontes de dados que possuem padrões distintos de registro e de atualização dos dados.

**Palavras-Chave:** Repositórios institucionais. Sistemas de informação. Interoperabilidade.

## ABSTRACT

This work addresses the interoperability of Lume, Digital Repository of Federal University of Rio Grande do Sul, with other information systems of UFRGS, which the goals are automate and improve its population with digital documents produced by University community, minimize inconsistencies, avoid the rework and data duplication, and keep the data integrity and uniformity. It describe the implementation of the Lume synchronization with data sources that have distinct patterns of records and data update.

**Keywords:** Institutional Repositories. Information Systems. Interoperability.

## Introdução

A interoperabilidade é um dos itens críticos quando se pensa no desenvolvimento e integração de repositórios e bibliotecas digitais. Utilizando o conceito de interoperabilidade as bibliotecas sempre trabalharam em torno da ideia do compartilhamento e da coope-

ração, tornando possível a integração de acervos digitais heterogêneos e distribuídos. Inicialmente os repositórios e as bibliotecas digitais eram estanques e as buscas deviam ser realizadas individualmente, utilizando interfaces distintas, e quando se desejava integrar as informações a solução, muitas vezes, era fazê-lo de forma manual. O objetivo sempre foi facilitar as buscas e tornar transparentes os processos realizados pelo usuário final.

Para Sayão e Marcondes (2008) a utilização de uma interface única e transparente para o usuário permite o acesso a toda a complexidade intrínseca das bibliotecas digitais e proporciona um mecanismo para mapear e integrar repositórios e bibliotecas digitais dando a impressão, para o usuário, de uma biblioteca única, onde a localização geográfica da informação deve ser indiferente.

A interoperabilidade é um tema central para tratar de coleções com uma grande variedade de tipos de dados, padrões de metadados, protocolos, esquemas de autenticação e modelos de negócios. Segundo Arms (2002), o objetivo da interoperabilidade é a construção de serviços coerentes para os usuários, a partir de componentes que são tecnicamente diferentes e de procedências diversas, o que requer acordos técnicos, de conteúdo e organizacionais. Os acordos técnicos referem-se a formatos, protocolos e sistemas de segurança, de modo que informações possam ser trocadas. Os acordos de conteúdo cobrem os dados e os metadados e incluem acordos semânticos sobre interpretação das informações. Por último, os acordos organizacionais referem-se às regras básicas para acesso, preservação de acervos e serviços, pagamento, autenticação, entre outros.

O desafio subjacente entre os gerentes de acervos, repositórios e bibliotecas digitais é que os participantes concordem em realizar pequenos esforços que permitam partilhar informações e serviços básicos, sem precisar abrir mão das características e identidade que tornam o conjunto de informações disponibilizadas únicas e essenciais para o usuário final.

O Centro de Processamento de Dados da Universidade Federal do Rio Grande do Sul (CPD-UFRGS) desenvolve e mantém sistemas de informação nas várias áreas de atuação da Universidade. Como resultado disso, eventualmente, verifica-se a dificuldade de obtenção de informações consolidadas e descrição de dados padronizada.

Os avanços proporcionados pelas tecnologias da informação e comunicação (TIC), no entanto, possibilitam a interoperabilidade entre os sistemas, de forma a evitar o retrabalho e a duplicação de dados, proporcionando sua consistência e uniformidade.

Integrar sistemas de informação distintos, mas que manipulam uma porção comum de dados, não é tarefa trivial. Por mais que trabalhem com um conjunto comum de dados, cada sistema é desenvolvido para solucionar um problema distinto e, portanto, sua construção e arquitetura são voltadas para uma perspectiva particular. Assim sendo, é imprescindível a adoção de normas e interfaces de comunicação padronizadas no desenvolvimento destes sistemas de informação, a fim de assegurar a comunicação transparente entre sistemas.

Este trabalho apresenta a estratégia e as ações implementadas para a interoperabilidade do Lume<sup>1</sup>, denominação atribuída ao repositório digital da Universidade Federal do Rio Grande do Sul (UFRGS), com alguns dos sistemas de informação da Universidade, considerando que esses sistemas armazenam informações cuja Universidade tem interesse



em disponibilizar em acesso aberto. A interoperabilidade é vista como um facilitador no processo de automatizar ao máximo o povoamento do Lume com documentos digitais produzidos na UFRGS, com o intuito de otimizar este processo, bem como garantir a completeza, consistência e integridade dos dados.

Está organizado da seguinte forma: a seção 2 aborda aspectos teóricos acerca de interoperabilidade e repositórios institucionais; a seção 3 apresenta o Lume, Repositório Digital da UFRGS; a seção 4 apresenta os sistemas de informação que são as fontes de dados para o Lume; a seção 5 relaciona os passos executados para a importação de dados dos vários sistemas da Universidade para o Lume e a seção 6 conclui o trabalho.

## **Interoperabilidade e repositórios institucionais**

É evidente a necessidade de partilhar a informação por meio de diferentes canais. Neste contexto, o estudo de Simões et al. (2014) destaca o diálogo entre os OPAC e os repositórios institucionais, possibilitando que o documento que se encontra depositado num ponto seja localizado e consultado em ambos. A interoperabilidade entre ambos efetiva o compartilhamento da informação e, deste modo, o papel que a própria informação e o conhecimento assumem nas Instituições do Ensino Superior, potencializando ainda mais o seu valor.

A interoperabilidade, no entanto, assume muitos outros aspectos que devem ser revistos e considerados, visto que o termo representa a ideia de comunicação entre entidades diferentes. Do ponto de vista técnico, trata-se da capacidade dos sistemas se comunicarem entre si. Mas, para além do aspeto técnico, é preciso que também intervenham políticas, para as quais relevam, entre outros fatores, os pactos e os protocolos estabelecidos para viabilizar esta comunicação. (SIMÕES et al., 2014). Portanto, deve ficar claro que há muito mais para garantir a interoperabilidade do que usar *hardware* e *software* compatíveis, embora sejam aspectos importantes. A garantia de interoperabilidade eficaz exige mudanças na maneira de organizar o trabalho e, especialmente, em relação às atitudes no registro e disponibilização das informações.

Certamente a face mais visível da interoperabilidade é, segundo Sayão e Marcondes (2008), a interoperabilidade técnica, por ser o aspecto mais perceptivelmente responsável por manter os sistemas de informação interoperáveis. Mais amplamente, a interoperabilidade pode ser considerada como o processo contínuo de assegurar que os sistemas, procedimentos e cultura de uma organização sejam geridos de forma a maximizar a troca e reutilização de informações. Para Miller (2000), Ukoln (2005) e Sayão e Marcondes (2008) ela abrange muitas áreas:

**a)** interoperabilidade técnica - inclui assegurar envolvimento de um conjunto de organizações no contínuo desenvolvimento de padrões de comunicação, transporte, armazenamento e representação de informações. Inclui também os esforços cooperativos para assegurar que padrões individuais evoluam em benefício da comunidade envolvida e para facilitar a convergência desses padrões, de forma que seja possível que os sistemas possam ter como base mais de um conjunto de padrões;

**b)** interoperabilidade semântica - está relacionada com o significado ou semântica das informações originadas de diferentes recursos e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da infor-

mação, como esquemas de metadados, classificações, tesouros e ontologias. Em muitos casos são usados termos diferentes para descrever conceitos semelhantes, como por exemplo: Autor, Criador e Compositor, ou mesmo utilizar termos idênticos para significar coisas muito diferentes, introduzindo confusão e erro no uso dos metadados;

**c)** interoperabilidade política/humana - diz respeito não apenas às questões relacionadas à maneira pela qual a informação é descrita e disseminada, a decisão de tornar os recursos informacionais mais amplamente disponíveis e interoperáveis tem implicações para a organização, para as equipes envolvidas e para os usuários em termos comportamentais, de recursos e de treinamento. A ênfase dada por parte de alguns setores governamentais aos problemas de democratização do acesso, da exclusão digital e da federação de fontes de informação voltadas para a educação a distância têm impacto nas políticas públicas para a área e estão enquadrados nesse item;

**d)** interoperabilidade intercomunitária - enfoca a necessidade impulsionada pela crescente interdisciplinaridade, principalmente nas áreas de pesquisa, de acesso a informações provenientes de um espectro amplo de fontes distribuídas por organizações, áreas de conhecimento e comunidades de natureza distintas. Geralmente exige o estabelecimento de fóruns para discussão e consenso em torno de práticas e procedimentos comuns;

**e)** interoperabilidade legal - considera as exigências e as implicações legais de tornar livremente disponíveis itens de informação;

**f)** interoperabilidade internacional - observa a diversidade de padrões e normas, os problemas de comunicação, as barreiras linguísticas, as diferenças no estilo de comunicação e a falta de uma fundamentação comum.

Considerando os aspectos acima, o CPD-UFRGS implantou o seu repositório digital, com o objetivo de reunir, preservar, divulgar e garantir o acesso confiável e permanente aos documentos científicos, técnicos e acadêmicos produzidos na Universidade, ampliando sua difusão e uso. Para viabilizar a interoperabilidade com outros sistemas de informação da Universidade, repositórios digitais, provedores de dados e/ou outros sistemas de informação externos à Universidade, desde o seu início, considerou-se a necessidade de adotar padrões e protocolos que permitissem a coleta de metadados e a transferência de informações.

Os repositórios institucionais, ao utilizarem metodologia de arquivos abertos (Open Archives Initiative), visam promover o acesso aberto à informação científica e acadêmica e, sobretudo, possibilitam armazenar, preservar e divulgar a produção intelectual, aumentando o impacto e a visibilidade das pesquisas desenvolvidas na instituição. (CAMARGO; VIDOTTI, 2006).

Do ponto de vista mais conceitual, constituem-se em autênticos sistemas de gestão de conteúdos, uma vez que além dos documentos em si, oferecem à comunidade acadêmica um conjunto de serviços para a gestão da produção institucional. (DIRETRIZES..., 2007). Destacam-se, dentre eles, as estatísticas de acessos e de downloads dos docu-

mentos, viabilizadas pelo uso do DSpace - Institutional Repository System<sup>2</sup>, ferramenta amplamente utilizada no desenvolvimento de repositórios institucionais, de acordo com o The Directory of Open Access Repositories - OpenDOAR<sup>3</sup>.

Os repositórios caracterizam-se, portanto, como ferramentas de gestão da informação e do conhecimento importantes não só no contexto da instituição, assim como contribuem com subsídios para as instituições de fomento à pesquisa.

## Sobre o Lume

O Lume adota, atualmente, a versão 1.8.1 do DSpace instalada em um servidor Linux. Esta foi a ferramenta escolhida por tratar-se de software livre que supre todas as propriedades que advêm das necessidades de divulgação das universidades, incluindo de forma nativa as seguintes características:

- a) facilidade de tradução e internacionalização;
- b) facilidade de customização e adequação às necessidades internas;
- c) utilização de um padrão de armazenamento de metadados universal e extremamente difundido e aceito, o *Dublin Core*;
- d) facilidade de indexação do seu conteúdo pelos principais sites de busca (p. ex., Google, Bing e etc.);
- e) compatibilidade com qualquer tipo de documento ou acervo que esteja em formato digital;
- f) navegação, pesquisa e consulta intuitiva, por meio de uma interface uniforme que não depende do usuário conhecer o conteúdo do repositório;
- g) exposição do conteúdo completo do repositório para *harvesting automatizado, mediante um protocolo de comunicação e consulta aberto, o Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*<sup>4</sup>.

Um repositório é composto por metadados e objetos digitais. No Lume, os metadados seguem o padrão Dublin Core qualificado - Dublin Core Metadata Initiative<sup>5</sup>.

Para garantir o acesso permanente aos documentos digitais, independente de qualquer mudança de endereço do servidor, utiliza-se o *Handle System*<sup>6</sup>, serviço fornecido pela *Corporation for National Research Initiatives* (CNRI), que consiste em atribuir identificadores persistentes para cada documento digital garantindo que, mesmo que o endereço do servidor do repositório digital mude, os recursos possam continuar sendo referenciados univocamente.

A interoperabilidade com outros repositórios digitais e/ou outros sistemas de informação externos à Universidade é viabilizada pelo uso do protocolo de coleta de metadados da iniciativa de arquivos abertos OAI-PMH, permitindo a transferência de dados entre eles.

<sup>1</sup>Lume, Repositório Digital da Universidade Federal do Rio Grande do Sul, <http://www.lume.ufrgs.br>

<sup>2</sup>DSpace Institutional Digital Repository System, <http://www.dspace.org>

<sup>3</sup>The Directory of Open Access Repositories - OpenDOAR, <http://www.opendoar.org/>

<sup>4</sup>Open Archives Initiative Protocol for Metadata Harvesting, <http://www.openarchives.org/pmh/>

<sup>5</sup>Dublin Core Metadata Initiative, <http://www.dublincore.org/documents/usageguide>

<sup>6</sup>The Handle System, <http://www.handle.net/>

## Sistemas de informação da UFRGS

A iniciativa de interoperar dados entre o Lume e outros sistemas da Universidade teve início quando da implantação do Repositório, em janeiro de 2008, com a inclusão das teses e dissertações pertencentes aos acervos do Sistema de Bibliotecas da UFRGS (SBUFRGS) e, portanto, devidamente descritas no Sistema de Automação de Bibliotecas (SABi).

Com a ampliação da abrangência do Lume outros sistemas passaram a alimentá-lo. Para aproveitar o processo de descrição, arquivamento e revisão das informações realizado nos sistemas fonte, a importação desses dados para o repositório institucional é feita de forma automatizada, por meio de extensões do DSpace.

Os sistemas que têm suas informações coletadas e disponibilizadas para a comunidade por meio do Lume são descritos a seguir.

### ***Sistema de Automação de Bibliotecas (SABi)***

O SABi realiza a gerência integrada das atividades e serviços das 32 bibliotecas e do catálogo on-line da UFRGS, por meio dos módulos de Aquisição, Catalogação, Itens, Periódicos e Circulação de Coleções. Adota padrões e normas internacionais para registro de dados bibliográficos, MARC 21<sup>7</sup>, e para intercâmbio de registros, ISO 2709<sup>8</sup> e ANSI/NISO Z39.2<sup>9</sup>. Utiliza o *software* Aleph 500 e o banco de dados Oracle, instalados no sistema operacional RedHat.

A Universidade, por intermédio do seu Sistema de Bibliotecas, tem priorizado, ao longo dos anos, a coleta, registro e disseminação da produção intelectual institucional (PI) no SABi. Os tipos de produção registrados na base de dados, atualmente, contemplam 48 categorizações, para os quais foram criados códigos, conforme a necessidade de identificá-los e contabilizá-los separadamente, manifestada pelas bibliotecas, pesquisadores e órgãos da Administração Central da Universidade.

A coleta da PI pelas bibliotecas e seu registro no SABi são prioridade no SBUFRGS. O trabalho sistemático das equipes neste sentido tem fortalecido e consolidado o SABi como instrumento de controle da PI e vem subsidiando a Administração Central na identificação dos indicadores de produtividade, os quais possibilitam dimensionar resultados mediante produtos acadêmicos produzidos no âmbito da Universidade.

Os dados relativos registrados no SABi são compartilhados com outros sistemas da Universidade, tais como: Avaliação institucional, Progressão docente, Relatório de Atividades Docentes, entre outros, além do Lume. Os registros de PI que povoam as comunidades do Lume são transferidos diariamente por meio de *harvesting* incremental.

O uso do SABi como fonte de informações de PI, integrado a diversas ferramentas e sistemas, tem incentivado a comunidade universitária a entregar, regularmente, suas produções nas respectivas bibliotecas para o devido registro na base de dados. No que concerne às equipes das bibliotecas, salienta-se a necessidade e importância da consistência dos registros na base, principalmente no que diz respeito à atribuição da autoria e classificação correta do tipo de produção, conforme as políticas e procedimentos do SBUFRGS relativos ao registro da PI da UFRGS, a fim de conferir confiabilidade ao produto final.

7 MARC Standards, <http://www.loc.gov/marc/>

8 ISO Standards, [http://www.iso.org/iso/iso\\_catalogue.htm](http://www.iso.org/iso/iso_catalogue.htm)

9 NISO Standards, [http://www.niso.org/apps/group\\_public/projects.php](http://www.niso.org/apps/group_public/projects.php)

### ***Sistema de Acervo Fotográfico***

É um sistema web para registro, armazenamento e consulta a arquivos fotográficos em formato JPG. Foi desenvolvido em PHP utilizando um banco de dados SQL Server, ambos instalados no sistema operacional Windows.

As fotos são descritas e armazenadas pelos respectivos órgãos/setores da Universidade responsáveis pelos acervos. O arquivamento é feito por meio do preenchimento de formulário com os dados da foto e do carregamento da imagem.

É importante destacar que a disponibilização das fotos em formato digital contribui muito para a preservação dos acervos físicos pois, desta forma, é possível evitar o manuseio do documento e, conseqüentemente, a sua deterioração. Este sistema é atualizado pelos órgãos da Universidade que o utilizam.

### ***Sistema de Eventos Institucionais (SEI)***

O desenvolvimento do Sistema de Eventos Institucionais teve início em 2009, por ocasião da organização do XXI Salão de Iniciação Científica (SIC), o qual serviu como “piloto” do sistema, abrangendo as etapas de submissão de trabalhos; a avaliação dos orientadores; a homologação das instituições externas; a inscrição de participantes; a organização das sessões de apresentação de trabalhos; a escolha de trabalhos premiados, entre outras. Até então, as informações referentes aos Salões de IC eram geridas por meio de um sistema específico para o Salão e que não servia aos demais eventos da Universidade.

O objetivo principal do SEI é disponibilizar um ambiente flexível, que possa ser usado para gestão dos diversos tipos de eventos promovidos pela Universidade como, por exemplo, conferências, seminários, congressos, salões, entre outros. Além disto, a disponibilização on-line dos trabalhos dos eventos tem evitado a publicação dos anais em papel.

O sistema foi desenvolvido em PHP, com vasto uso de orientação a objetos, usando o framework Yii. A base de dados do SEI é o servidor SQL Server que dá suporte à base de dados institucional da Universidade, o que permite acesso direto a todas as informações de suporte que podem ser necessárias nos diversos tipos de eventos que o sistema irá gerir. Tanto o portal em PHP como o banco de dados SQL Server estão instalados no sistema operacional Windows. Além dos metadados relativos aos eventos como um todo e aos trabalhos específicos de cada evento, os arquivos originais dos autores (pôsteres, resumos, etc.) também são mantidos.

Em 2010 foi iniciado o projeto de criação do Portal de Eventos Institucionais, um ambiente integrado que permite ao usuário acessar todos os serviços que estão à sua disposição. O sistema é responsável por identificar o perfil do usuário em cada evento disponível no sistema (participante, orientador, avaliador de trabalhos, organizador, gerente, etc.) e permitir o acesso aos serviços para o respectivo perfil.

A solução adotada para o SEI não é desenvolver um sistema genérico, no sentido de que deva prever todos os casos, e sim um sistema adaptável. Neste aspecto, o mecanismo de serviços e perfis é de extrema utilidade, permitindo a criação de novos serviços,



ao invés da modificação dos já existentes, o que mantém a complexidade do sistema baixa e facilita sua adequação para novas situações. O mecanismo em que o sistema de adaptação se baseia é também bastante simples e depende apenas da execução de consultas SQL para liberar serviços, associar serviços a perfis, associar perfis a pessoas e assim por diante. (CANTO et al., 2011).

Neste sistema a atualização das informações é feita desde o cadastramento até o fechamento do evento no sistema. Depois do fechamento do evento, suas informações não são mais modificadas. A partir deste momento todos os trabalhos apresentados serão transferidos automaticamente para o Lume povoando a comunidade Eventos UFRGS.

### **Sistema de Legislação**

O Sistema de Legislação da UFRGS foi criado com o objetivo de cadastrar e disponibilizar as decisões do Conselho Universitário (CONSUN) e as resoluções do Conselho de Ensino Pesquisa e Extensão (CEPE). É um sistema *web* desenvolvido em PHP, utilizando o framework Yii e banco de dados SQL Server.

As decisões e resoluções cadastradas no sistema podem ter três tipos de relacionamento entre si: alterar, citar e revogar. Cada uma tem o respectivo arquivo PDF, escaneado a partir da versão impressa, com reconhecimento OCR. Além disso, é possível incluir um arquivo com a versão consolidada de uma decisão ou resolução que tenha sido alterada. O sistema é atualizado pelo CEPE e CONSUN.

No Quadro 1 são relacionadas as comunidades/subcomunidades/coleções que integram o Lume, com a indicação da respectiva fonte dos dados e o total de objetos digitais incluídos a partir do Sistema de Automação de Bibliotecas (SABi), do Sistema de Eventos Institucionais (SEI), do Sistema de Acervo Fotográfico e do Sistema de Legislação.

**Quadro 1. Fonte dos dados que alimentam as comunidades, subcomunidades e coleções do Lume.**

<b>Comunidades</b>	<b>Subcomunidades/ Coleções</b>	<b>Fonte dos dados</b>	<b>Depósitos*</b>
Acervos	CEME - Centro de Memória do Esporte	Auto-arquivamento e SABi	6.695
	Instituto de Física	Sistema Acervo Fotográfico	258
	Legislação UFRGS	Sistema de Legislação	4.075
	Museu Universitário	Sistema Acervo Fotográfico	474
	Setor de Patrimônio Histórico	Sistema Acervo Fotográfico	2.871
	TelessaúdeRS	SABi	127



Comunidades	Subcomunidades/ Coleções	Fonte dos dados	Depósitos*
Eventos UFRGS	Feira de Ensino e Popularização da Ciência - FEPOP	SEI	179
	Feira de Iniciação Científica	SEI e SABi	366
	Feira de Inovação Tecnológica - FINOVA	SEI	470
	Salão de Ensino	SEI	2.959
	Salão de Extensão	SEI	671
	Salão de Iniciação Científica	SEI e SABi	49.946
Trabalhos acadêmicos e técnicos	Apostilas	SABi	52
	Relatórios Técnicos e de Pesquisa		78
	Trabalhos de Conclusão de Curso de Especialização		4.720
	Trabalhos de Conclusão de Curso de Graduação		15.049
	Traduções		82
Produção intelectual	Artigos de Periódicos	SABi	14.382
	Entrevistas		537
	Livros e Capítulos de Livros		453
	Propriedade Intelectual - Patentes		239
	Resenhas		127
	Trabalhos de Eventos		28.151
Teses e dissertações	Teses e Dissertações defendidas na UFRGS	SABi	
		26.583	
	Teses e Dissertações não defendidas na UFRGS		251

\* *Dados extraídos do Lume em 28/07/2016.*

## Importação de dados e interoperabilidade

A importação dos dados dos sistemas citados na seção anterior é realizada de forma diferenciada, levando em conta a periodicidade com que os mesmos são incluídos. No caso do SABi, do Sistema de Acervo Fotográfico e do Sistema de Legislação a importação dos dados é realizada diariamente, permitindo inclusive a atualização daqueles que já foram importados anteriormente. No caso do SEI, a importação é feita por demanda, normalmente ao final de cada evento que se deseja disponibilizar.

O processo de importação dos dados é realizada por meio de *scripts* PHP que consultam as informações nos diferentes sistemas e, em etapa subsequente, os seguintes passos são executados:

**1)** criação de log do processo de importação que permite salvar o andamento da importação para detectar possíveis erros ocorridos durante o mesmo;

**2)** criação de *links* na pasta *assetstore* - neste diretório o DSpace salva os arquivos e é neste diretório que ele espera encontrar os arquivos, então foram feitos *links* simbólicos do Linux dentro da pasta *assetstore* que fazem referência aos arquivos nos sistemas fonte, evitando a cópia dos arquivos;

**3)** execução do processo de carga dos metadados do banco de dados do sistema fonte para o banco de dados do Lume, convertendo os metadados para o formato Dublin Core. Para inserir um item no Lume com seus respectivos metadados e referências aos arquivos são acessadas as seguintes tabelas do banco de dados do DSpace, nessa ordem:

- I. public.item
- II. public.handle
- III. public.collection2item
- IV. public.bitstream
- V. public.bundle
- VI. public.item2bundle
- VII. public.bundle2bitstream
- VIII. public.metadatavalue
- IX. public.resourcepolicy

Neste ponto são executados quatro comandos fornecidos pelo DSpace e localizados dentro da pasta bin do diretório de instalação da respectiva ferramenta. São esses os passos 4, 5, 6 e 7.

**4)** execução do comando ***cleanup*** para limpar arquivos antigos;

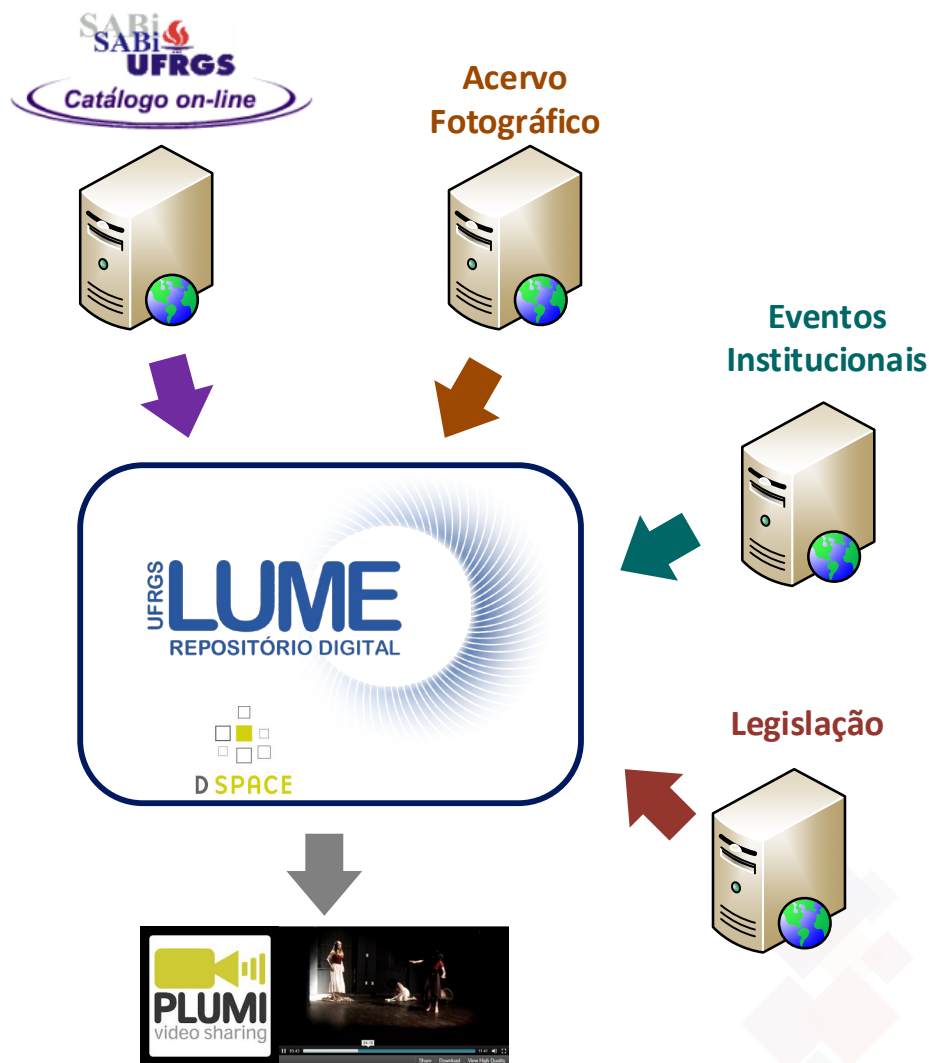
**5)** execução do comando ***filter-media*** para criar as miniaturas (thumbnails) das imagens e PDFs que irão aparecer na busca de registros;

**6)** execução do comando ***index-update*** para atualizar os índices de pesquisa para os novos itens;

**7)** execução do script ***sub-daily***, que envia os e-mails com os novos itens das coleções para os respectivos assinantes e, por fim,

**8)** envio do arquivo de log para a gerência do Lume, caso ocorra algum erro em um dos passos da importação de dados.

A Figura 1 ilustra os sistemas que interoperam com o Lume.



**Figura 1 - Visualização da interoperabilidade do Lume com os sistemas de informação da UFRGS.**

No diagrama estão representados os vários sistemas que alimentam o Lume e a relação entre eles.

Com a popularização da utilização de recursos audiovisuais para complementar e enriquecer trabalhos, além da própria necessidade do arquivamento e divulgação de vídeos e áudios relacionados a eventos e a teses e dissertações, por exemplo, o Lume conta cada vez mais com itens que contêm arquivos audiovisuais.

Tendo em vista o crescimento deste tipo de material identificado nos documentos incluídos no Lume e a necessidade de otimizar o acesso aos mesmos, foi adotado o servidor de streaming open source Plumi<sup>10</sup>, o qual permite o streaming de vídeos do repositório, não demandando o download completo do arquivo nem de um programa adicional. Esta ferramenta tem se mostrado capaz de reproduzir diversos formatos de vídeo para praticamente todos os dispositivos com acesso à internet existentes, tais como, browsers, tablets e telefones. Atualmente a inclusão de vídeos do Lume no Plumi é realizada de forma manual e está em estudo uma forma de automatizá-la.

10 Plumi video sharing, <https://plone.org/products/plumi>

## Considerações finais

As soluções que vêm sendo gradualmente adotadas com relação à interoperabilidade do Lume com os sistemas de informação da Universidade consolidam a recomendação explícita no Plano de Desenvolvimento de Tecnologia de Informação (PDTI), elaborado pela Universidade para o período de 2011 a 2015, sobre a necessidade da adoção de soluções institucionais que permitam a disponibilização das informações geradas pelos diversos órgãos gestores da UFRGS de forma integrada, a fim de evitar a redundância de dados e a proliferação de informações sem a garantia de sua procedência, confiabilidade e responsabilidade. (UNIVERSIDADE ..., 2011).

A integração e o compartilhamento de dados entre os diversos sistemas têm se mostrado extremamente positivos, sobretudo pelo aproveitamento dos metadados, evitando o retrabalho e reduzindo a possibilidade de erros e/ou inconsistência dos mesmos. A manutenção do dado em um único sistema é fundamental como meio de resguardar sua consistência e veracidade.

A colaboração interinstitucional também tem sido discutida no contexto da interoperabilidade e graças à adoção de padrões internacionais as informações do Lume são coletadas e estão disponíveis em diversos sistemas externos como a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)<sup>11</sup> e o Oasisbr<sup>12</sup>, ambos do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), a Red Federada de Repositorios Institucionales de Publicaciones Científicas - LA Referencia<sup>13</sup>, o Repositório Científico de Acesso Aberto de Portugal (RCAAP)<sup>14</sup>, a Networked Digital Library of Theses and Dissertations (NDLTD)<sup>15</sup>, entre outros.

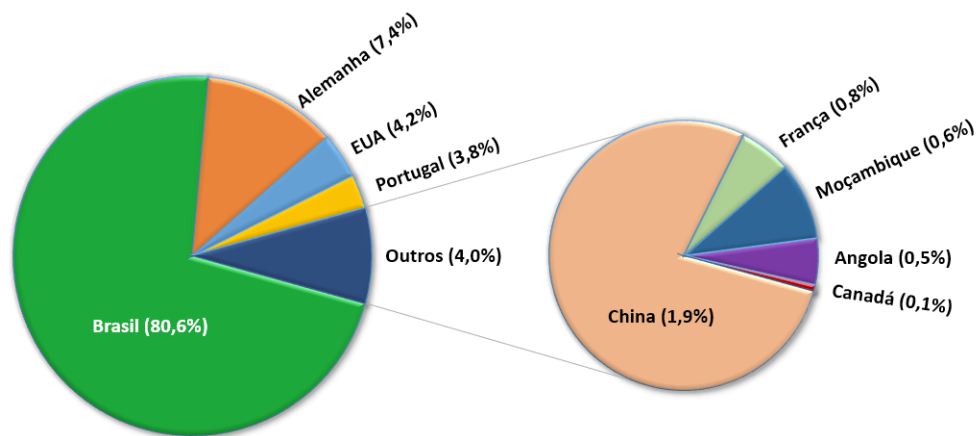
De acordo com Arellano (2008) é nessa estrutura que estão apoiados todos os aspectos dos repositórios digitais de acesso livre. Estrategicamente, essa abertura e cooperação geram uma multiplicação de benefícios para todos os participantes, incluindo novas habilidades e experiências, modelos de negócios sustentáveis, estratégias de preservação mais estruturadas, entre outros. O uso de padrões abertos e de protocolos para a interoperabilidade dos arquivos permite a incorporação de colaboradores internacionais, formalizando redes de informação que minimizam a redundância dos recursos e custos na sua manutenção.

Pode-se afirmar que a integração do Lume com os sistemas de informação existentes na Universidade e sistemas externos privilegia não apenas a interoperabilidade técnica no desenvolvimento de padrões de comunicação, de transferência de dados e de armazenamento, como também a interoperabilidade semântica, incentivando o uso de padrões terminológicos e conceituais consistentes e que representem o conteúdo informacional de forma precisa; a interoperabilidade política/humana tornando o Lume um instrumento institucional de difusão da produção científica e de democratização da informação, que para seu sucesso envolve pessoas das diversas instâncias da Universidade; a interoperabilidade intercomunitária, disponibilizando informações das diversas áreas de pesquisa da Universidade, ampliando o acesso à informação de interesse geral da comunidade científica a partir de uma ampla gama de fontes e Comunidades disponibilizadas; a interoperabilidade legal, resguardando os direitos dos autores e implementando licenças que regulam o uso da informação e a interoperabilidade internacional, minimizando as variáveis nos padrões, problemas de comunicação e as barreiras linguísticas. Esta última com a adoção de metadados em outros idiomas considerando a diversidade de

usuários que acessam o Repositório.

A interoperabilidade entre diferentes sistemas atualmente é, acima de tudo, uma necessidade, permite a comunicação direta e transparente com informações corretas e instantâneas, a fim de aperfeiçoar esforços para disponibilizar a quantidade de informações, imprescindíveis para a gestão e visibilidade das Instituições de Ensino Superior e tornar públicos os resultados das pesquisas desenvolvidas na Universidade, contribuindo para a socialização do conhecimento. O alcance nacional e internacional do Lume pode ser observado no Gráfico 1, que mostra o percentual por país, num total de 67.089.162 downloads, no período de 2008 a julho de 2016.

Gráfico 1 - Percentual de downloads por país, 2008-2016



*Dados extraídos do Lume em 29/07/2016.*

11 <http://bdtd.ibict.br/vufind/>  
12 <http://oasisbr.ibict.br/vufind/>  
13 <http://lreferencia.redclara.net/rfr/>  
14 <https://www.rcaap.pt/>  
15 <http://www.ndltd.org/>

## Referências

Arellano, M. A. (2008). Critérios para a preservação digital da informação científica. Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, DF, Brasil. Recuperado em 6 junho, 2016, de <http://repositorio.unb.br/handle/10482/1518>

Arms, W. Y., Hillmann, D., Lagozes, C., Krasst, D., marisa, R., Saylor, J., Terrizzi, C. & Van de Sompel, H. (2002). A spectrum of interoperability: the site for science for prototype for the NSDL [Versão eletrônica], D-Lib Magazine, v. 8, n. 1. Recuperado em 3 junho, 2016, de <http://www.dlib.org/dlib/january02/arms/01arms.html>

Camargo, L. S. A. & Vidotti, S. A. B. G. (2006). Elementos de personalização em repositórios institucionais. Trabalho apresentado na 1ª Conferência Iberoamericana de Publicações Eletrônicas no Contexto da Comunicação Científica, Brasília.

Canto, F. H., Machado, J. L., Cortinovi, S. L., Neves, R. M. das, Ferreira, M. K., Closs, U. C., Araújo, A. M., Lopes, J. R., Benites, M. R. & Castaldello, C. J. B. (2011). Sistema de Eventos Institucionais. Trabalho apresentado no 5º Workshop de Tecnologia da Informação das Instituições Federais de Ensino Superior, Florianópolis. Recuperado em 15 abril, 2016, de <http://www.vwticifes.ufba.br/modulos/submissao/Upload/33171.pdf>

Diretrizes para a criação dos repositórios institucionais nas universidades e organizações de educação superior. (2007). Pontifícia Universidad Católica de Valparaíso, Valparaíso. Recuperado em 30 março, 2016, de [http://eprints.rclis.org/bitstream/10760/13512/3/Diretrizes\\_RI\\_portugues.pdf](http://eprints.rclis.org/bitstream/10760/13512/3/Diretrizes_RI_portugues.pdf)

Miller, P. (2000). Interoperability. What is it and why should I want it? Ariadne, n. 24. Recuperado em 8 junho, 2016, de <http://www.ariadne.ac.uk/issue24/interoperability/>

Sayão, L. F. & Marcondes, C. H. (2008). O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais [Versão eletrônica], Transinformação, v. 20, n. 2. Recuperado em 3 maio, 2016, de <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/530>

Simões, M. G. M., M., Freitas, C. V. de, Bravo, B. R., & Gomes, S. (2014). OPAC e repositórios institucionais: uma análise preliminar sob a perspectiva da interoperabilidade [Versão eletrônica], Cadernos BAD, n. 2. Recuperado em 3 maio, 2016, de <http://bad.pt/publicacoes/index.php/cadernos/article/download/1205/1212>

Ukoln. (2005) Interoperability focus: looking at interoperability. Recuperado em 8 junho, 2016, de <http://www.ukoln.ac.uk/interop-focus/about/leaflet.html>

Universidade Federal do Rio Grande do Sul. (2011). Plano de Desenvolvimento de Tecnologia de Informação (PDTI): relatório final 2011-2015. Recuperado em 28 março, 2016, <http://www.ufrgs.br/ufrgs/a-ufrgs/pdti-plano-diretor-de-tecnologia-da-informacao>

# “Desarrollo de un sistema de gestión de investigación (CRIS) para la consolidación del Repositorio Institucional NINIVE de la Universidad Autónoma de San Luis Potosí”

*Autores: MTE Rosalina Vázquez Tapia; Ing. Claudia Lizett Amaya Varela.; L.B. Elizabeth Elena Hernández Peña ; Josué Martínez Vázquez.*

---

## Resumen

La Biblioteca Virtual Universitaria (BVU) de la Universidad Autónoma de San Luis Potosí (UASLP) fue creada en el 2006 bajo un modelo propio como parte de la estructura del Sistema de Bibliotecas. El Repositorio Institucional NINIVE, implementado por la BVU en el 2012 contiene a la fecha (agosto de 2016) más de 2,450 documentos digitales entre tesis de licenciatura y posgrado, libros, artículos de revistas universitarias, objetos de aprendizaje e informes de Rectoría.

Para la consolidación del Repositorio Institucional NINIVE en una primera fase, se presentó un proyecto para su financiamiento en la Convocatoria pública 2015 del Consejo Nacional de Ciencia y Tecnología (CONACYT), mismo que fue aprobado para su ejecución durante el periodo junio – noviembre de 2016.

El proyecto consiste en desarrollar un Modelo de Servicios para la consolidación del Repositorio Institucional NINIVE de la UASLP. Uno de sus objetivos o componentes principales es el desarrollo a nivel prototipo de un Sistema Institucional de Gestión de Investigación (SIGI) bajo un modelo de CRIS (*Current Research Information System*), como una estrategia de sensibilización y promoción con la comunidad científica para el depósito o auto archivo de las publicaciones y productos de investigación en el Repositorio NINIVE.

Para la implementación del sistema se proponen cinco estrategias generales: 1) Realizar un estudio de benchmarking; 2) Diseñar el currículum universitario u hoja de vida del investigador; 3) Adoptar el estándar internacional ORCID para generar los identificadores únicos digitales de los investigadores; 4) Diseñar e implementar los módulos del sistema y aplicaciones (API's) necesarias para su interoperabilidad; 5) Integrar un grupo piloto de investigadores para fines de retroalimentación, validación y pruebas de usabilidad.

El objetivo de esta propuesta es presentar el modelo y arquitectura del sistema SIGI así como también los resultados y avances del proyecto.

## Palabras clave:

**Repositorios Institucionales, Acceso Abierto, Interoperabilidad, Sistema de gestión de investigación (CRIS)**



## **Abstract**

The Virtual Library (BVU) of the Autonomous University of San Luis Potosi (UASLP) was created in 2006 under its own model as part of the structure of the Library System. The Institutional Repository NINIVE was implemented by the BVU in 2012. Actually (August 2016) it contains more than 2,450 digital documents between undergraduate and graduate theses, books, articles of academic journals, learning objects and Reports form Rectory.

To consolidate the Institutional Repository NINIVE in a first phase, a project was presented for funding in the Public Call 2015 of the National Council of Science and Technology Information (CONACYT), which it was approved for implementation during the period June - November 2016.

The project's goal is to develop a Services Model for consolidation of UASLP Institutional Repository. One of its aims or components is the development of a Research Management Institutional System to prototype level, under a Current Research Information System (CRIS) model, as a strategy for awareness raising and promoting to the scientific community for the deposit or archive self of publications and research products in the NINIVE repository.

To implement the system five broad strategies are proposed: 1) Conduct a benchmarking study; 2) Design the academic curriculum or researcher profile; 3) Adopt the ORCID international standard to generate unique digital identifiers of researchers; 4) Design and implement the system modules and applications (APIs) necessary for interoperability; 5) Integrate a pilot researchers group for feedback purposes, validation and usability testing.

The objective of this proposal is to present the model and SIGI system architecture as well as the results and progress of the project.

## **Keywords**

**Institutional Repositories, Open Access, Interoperability, Current Research Information System (CRIS)**

## **Antecedentes**

La Biblioteca Virtual de la Universidad Autónoma de San Luis Potosí (UASLP) fue creada en el 2006 como parte de la estructura del Sistema de Bibliotecas, con el propósito fundamental de ofrecer a la comunidad universitaria y a la sociedad, el acceso a colecciones digitales universitarias, recursos electrónicos de información y servicios de tecnología educativa, para apoyar el desarrollo de las funciones sustantivas de la Institución.

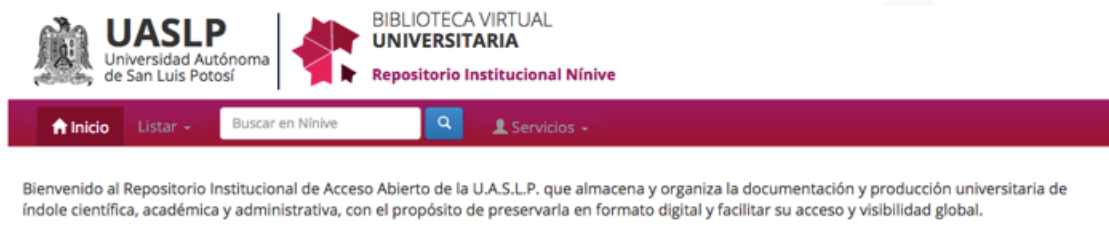
Para el desarrollo de la Biblioteca Virtual se diseñó un modelo propio basado en cuatro principios: La integración de tres tipos de acervos (suscritos, abiertos y propios), el desarrollo de servicios virtuales personalizados, la utilización de estándares y tecnologías abiertas y la vinculación con la academia.

Con base en este modelo, en el 2008 se inició un programa de digitalización de las tesis de licenciatura y posgrado disponibles en las diferentes unidades de información del Sistema de Bibliotecas. Las tesis fueron digitalizadas con el apoyo de estudiantes de servicio social y catalogadas por el personal bibliotecario utilizando estructuras de metadatos en el estándar *Dublin Core*, de tal manera que las colecciones digitales pudieran ser migradas posteriormente a una plataforma de gestión de repositorios.

Hasta hace dos años, el acceso y consulta a las tesis digitales junto con otras colecciones universitarias como libros publicados por la Editorial Universitaria y revistas de divulgación, era a través del portal principal de la Biblioteca Virtual denominado CREATIVA (Centro de Recursos Académicos, Informáticos y Virtuales)<sup>1</sup>, donde se integran los recursos electrónicos de suscritos por la UASLP, los disponibles a través del Consorcio Nacional de Recursos de Información Científica y Tecnológica (CONRICYT) y una selección de recursos de Acceso Abierto.

En 2012 se inició el proceso de planeación, diseño y construcción del Repositorio Institucional de la U.A.S.L.P. al que se le dio el nombre de NINIVE. Este proceso consistió en el desarrollo de cinco acciones estratégicas: 1) Instalación y personalización del software *DSpace* de código abierto para la gestión del repositorio en un servidor dedicado; 2) Diseño de la estructura del repositorio con base en los tipos de producción o acervo (comunidades), las dependencias y programas académicos (sub comunidades) y tipos de contenidos (colecciones); 3) Migración de las colecciones digitales universitarias de CREATIVA a NINIVE; 4) Diseño del flujo de trabajo, los lineamientos para la gestión de derechos de autor basados en el uso de licencias abiertas y las políticas de servicio; 5) Diseño de un Plan de difusión y capacitación del personal administrativo y académico para el auto archivo y gestión de colecciones en el Repositorio. Un año después, entró en operación el Repositorio Institucional NINIVE<sup>2</sup> el cual contiene actualmente (agosto 2016) 2148 documentos digitales entre tesis, libros, artículos de revistas universitarias, objetos de aprendizaje e informes de Rectoría. Véase Figura 1.

Los servicios asociados a NINIVE son esencialmente la digitalización de documentos de diversos tipos (impresos, fotografías, videos), la capacitación al personal de gestión y a los investigadores y el soporte técnico.



1 Centro de Recursos Académicos, Informáticos y Virtuales. <http://creativa.uaslp.mx>

2 Repositorio Institucional NINIVE de la UASLP. <http://ninive.uaslp.mx>

### Comunidades en Nínive

Elija una comunidad para visualizar sus colecciones.

- Archivo Institucional**  
Documentos administrativos y de gestión
- Documentación Científica**  
Publicaciones arbitradas y productos de investigación
- Patrimonio documental**  
Fondo antiguo y colecciones históricas
- Publicaciones y Producciones Universitarias**  
Materiales editoriales, audiovisuales, fotográficos y multimedia
- Recursos Educativos**  
Materiales didácticos, trabajos escolares y documentos académicos
- Registros bibliográficos de CREATIVA**  
Depósito de intercambio bibliográfico de las colecciones universitarias actualmente disponibles en CREATIVA

### Consulta

Autor		Tema		Fecha de publicación	
García Valdez, Mario	9	ovejas	9	2000 - 2016	1267
Lastras Ramirez, Alfonso	8	Suelos	9	1900 - 1999	174
Melendez de Espinosa, Juana	8	Cuencas hidrográficas	7	1554 - 1599	2
Valle Méndez, Jaime	8	Chile	6		
Montejano y Aguiñaga, Rafael	7	Cultivo	6		
Margain Manautou, Emilio	6	Coloides	5		
Instituto de Ciencias Educativas	5	Corderos	5		
Meléndez, Juana	5	tomate	5		
Padilla Segura, José Antonio	5	tomates	5		
Ramírez Delgado, Juan Manuel	5	Arquitectura	4		
<a href="#">siguiente &gt;</a>					

**Figura 2. Sitio principal del Repositorio Institucional NINIVE de la U.A.S.L.P.**

Adicionalmente, la BVU ofrece a la comunidad universitaria los servicios de la Red Universitaria de Videoconferencia y la plataforma educativa TZALOA para la formación de usuarios en línea. A nivel de colaboración interinstitucional, la Dirección de BVU tiene a su cargo la Coordinación General de la *Red Mexicana de Repositorios Institucionales – REMERI*<sup>3</sup> y es Responsable Técnico en la *Red Federada de Repositorios Institucionales de Publicaciones Científicas LA-REFERENCIA*<sup>4</sup>.

## Introducción

Uno de los aspectos clave para lograr los propósitos de difusión, acceso y preservación de un Repositorio Institucional (RI) es la implementación de un modelo de servicios, que se refiere a los servicios que ofrece un repositorio para promover el auto archivo, mantener el crecimiento o poblamiento a largo plazo y asegurar su sostenibilidad operativa y financiera.

De acuerdo a Swan (2008), los modelos de negocio para repositorios bajo el enfoque de servicio, operan bajo tres componentes principales; uno es la viabilidad, que implica considerar las necesidades y preferencias de los usuarios, definir los objetivos de la institución, los tipos de materiales, las políticas y el flujo de trabajo; el otro es la sostenibilidad, que se refiere a los recursos y costos de configuración e implementación; y el tercero es la adaptabilidad, concerniente a la flexibilidad, resistencia, tendencias y nuevos desarrollos.

Con respecto a los servicios que brindan los repositorios, existe una amplia gama de ellos. Éstos son implementados por las organizaciones o instituciones y se adoptan según sus necesidades y sus recursos. Algunos ejemplos de servicios son: la digitaliza-

<sup>3</sup> Red Mexicana de Repositorios Institucionales. <http://www.remeri.org.mx>

<sup>4</sup> LA-REFERENCIA. <http://lareferencia.redclara.net/rfr/>

ción de documentos; la asesoría sobre licencias o copyright; los servicios de *hosting*; la capacitación y el soporte técnico; los avisos y noticias a través de las redes sociales; las estadísticas; la preservación digital; el análisis de datos y metadatos; las revistas institucionales y los servicios de publicación; la vinculación con herramientas de Web social y métricas alternativas; y el desarrollo e integración con sistemas CRIS (*Current Research Information System*).

Un Sistema de Gestión de la Investigación o CRIS (por sus siglas en inglés) es una herramienta que permite gestionar de manera global todos los procesos relacionados con la investigación de una institución. Son sistemas que facilitan la gestión, recuperación y enriquecimiento de la información y producción científica de una institución que al vincularse a los repositorios institucionales y a otras aplicaciones, permiten proveer información estadística e indicadores de producción científica de una manera estandarizada e integral.

Los CRISs recolectan un amplio rango de metadatos relacionados con todos los aspectos de las actividades de investigación realizadas por una institución. Estos sistemas han sido desarrollados para asistir a los usuarios en el registro, generación de información y toma de decisiones concerniente al proceso de investigación, ya sea que ellos desarrollen programas, ofrezcan financiamiento, asesoren o ejecuten proyectos, generen resultados o transfieran tecnología (EroCRIS, 2016).

En los años recientes se ha registrado un crecimiento notable en la adopción de CRIS y sistemas de repositorios en las universidades alrededor del mundo. Según Castro, Shearer & Summan(2014), en la actualidad los RIs se han enfocado principalmente a la recolección de artículos de investigación de igual forma que las tesis y disertaciones, observando un interés creciente por expandir su alcance a la recolección de datos de investigación; desde la perspectiva del CRIS, las publicaciones son el resultado de proyectos relacionados con actividades institucionales mientras que desde el punto de visto del RI las publicaciones son recursos académicos que deben estar disponibles para su uso.

De acuerdo a las recomendaciones que hace EuroCRIS (2016), para el desarrollo de un CRIS se requiere primeramente de una plataforma *open source* interoperable como *DSpace-CRIS*<sup>5</sup> para recolectar, gestionar, preservar y diseminar información relacionada con entes de investigación tales como recursos (investigadores y sus intereses, referencias, posiciones, unidades organizacionales, grupos de investigación, etc.); actividades (proyectos, convenios, colaboraciones, etc.); y productos (publicaciones, reportes, patentes, etc.). Estas entidades y sus atributos y relaciones son modeladas de acuerdo a ontologías tales como *CERIF* que es un estándar mundial para el modelado de dominio de información de investigación.

En segundo término, es necesario adoptar identificadores únicos persistentes para los datos de investigación (DOIs) y para los investigadores (ID). *DataCite*<sup>6</sup> es una organización global sin fines de lucro que soporta la creación de DOIs y ofrece a la comunidad de investigación un conjunto de servicios de acuerdo a sus necesidades y promueve una cultura abierta de datos compartidos. Por su parte, *ORCID*<sup>7</sup> es una organización sin fines de lucro que trabaja con comunidades de investigación integradas por investigadores, universidades, editoriales, agencias de financiamiento, repositorios de datos y asociaciones internacionales. *ORCID* resuelve la ambigüedad de nombres de autores y les permite obtener el crédito por su trabajo a través de los sistemas integrados con registro *ORCID*. Permite conexiones transparentes y confiables entre investigadores, sus contribuciones y afiliaciones, simplificando el análisis y obtención de información.

5 DSpace-CRIS. <https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home>

6 DataCite. <https://www.datacite.org/>

7 ORCID. <https://orcid.org/>

Por último, la interoperabilidad entre las diferentes plataformas se logra a través de protocolos de intercambio de información tales como OIA-PMH, REST y SWORD, así como la implementación de APIs.

## Justificación y desarrollo del proyecto

La UASLP es una Institución Pública de Educación Superior que ha recibido ocho Reconocimientos Nacionales de Calidad<sup>8</sup>. Actualmente ofrece 84 programas educativos de posgrado y 92 de licenciatura; cuenta con una planta académica de poco más de 3,000 profesores de los cuales 787 son de tiempo completo (43% en el Sistema Nacional de Investigadores); atiende a una población estudiantil de 28,376 alumnos de licenciatura y posgrado (matricula al 31/diciembre/2014)<sup>9</sup>.

En cuanto a indicadores institucionales de producción de recursos de información científica, tecnológica y de innovación, de acuerdo a los resultados reportados en los Informes de Rectoría de los últimos cuatro años, se generan un promedio de 2,500 productos de investigación entre artículos científicos arbitrados, libros, tesis de posgrado y otros documentos. A nivel institucional, en los últimos cinco años se han publicado poco más de 3,000 artículos científicos arbitrados en revistas indexadas en los sistemas *Web of Science* y *Scopus*.

De esta producción sólo están disponibles en NINIVE un 40% de las tesis de posgrado y un 2% de otros productos de investigación. Como una estrategia de poblamiento y sostenibilidad para la consolidación del Repositorio a largo plazo, se presentó el proyecto titulado “Desarrollo de un Modelo de Servicios para la consolidación del Repositorio Institucional NINIVE de la U.A.S.L.P.” en la *Convocatoria 2015 para desarrollar los repositorios institucionales de acceso abierto a la información científica, tecnológica y de innovación del Consejo Nacional de Ciencia y Tecnología (CONACYT)*, mismo que fue aprobado para su financiación y ejecución durante el periodo julio – diciembre de 2016.

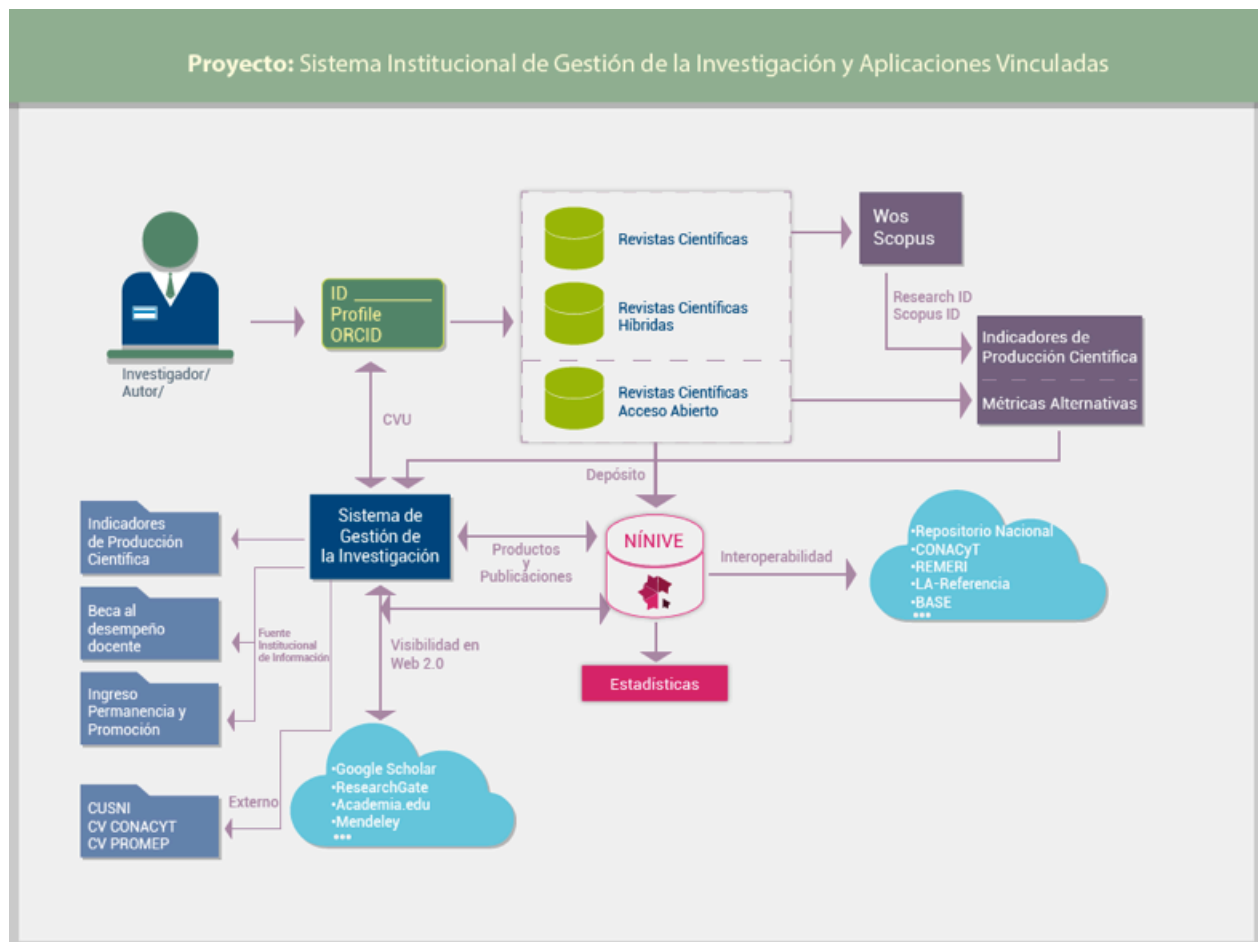
El proyecto consiste en desarrollar un Modelo de Servicios para la consolidación del Repositorio Institucional NINIVE de la UASLP, que comprende cinco objetivos estratégicos: **1)** El desarrollo a nivel prototipo de un Sistema de Gestión de Investigación interoperable con NINIVE como una forma de incentivar a la comunidad científica al auto archivo y publicación científica en AA; **2)** La actualización de la plataforma tecnológica para asegurar la interoperabilidad con OpenAIRE; **3)** La digitalización de recursos de información científica y tecnológica para su depósito en el repositorio; **4)** La preservación digital de los documentos para su reutilización a largo plazo; **5)** El diseño de un Mandato Institucional para el auto archivo de la producción científica, académica y documental universitaria.

Para la implementación del Sistema Institucional de Gestión de la Investigación (SIGI) se proponen cinco estrategias generales: 1) Realizar un estudio de benchmarking para identificar las mejores prácticas y aplicaciones desarrolladas por otras instituciones para la implementación de sistemas CRIS; 2) Diseñar el currículum universitario u hoja del vida del investigador; 3) Adoptar el estándar internacional ORCID para generar los identificadores únicos digitales de los investigadores; 4) Diseñar e implementar los módulos del sistema y aplicaciones (API's) necesarias para su interoperabilidad; 5) Integrar un grupo piloto de investigadores de diferentes áreas de conocimiento la UASLP para fines de retroalimentación, validación y pruebas de usabilidad.

8 En: <http://www.uaslp.mx/Paginas/Universidad/Informes/Indicadores/2015/Premios-y-reconocimientos.aspx>



El modelo del sistema SIGI para la UASLP consiste primeramente en diseñar el currículum universitario u hoja de vida del investigador con base en tres perfiles: El interno para evaluación y desempeño docente, el externo que contempla los requisitos del Sistema Nacional de Investigadores (SNI) y el Programa para el Desarrollo Profesional Docente (PRODEP) y el internacional que considera el estándar ORCID para generar los identificadores únicos digitales de los investigadores. Con la implementación de interfaces de aplicaciones (API's) de ORCID será posible hacer consultas por investigador a las bases de datos como *Scopus* y *Web of Science* y a la mayoría de las editoriales que ya forman parte de esta iniciativa, exportando los resultados hacia el sistema SIGI y hacia NINIVE, de tal forma que si un investigador actualiza su perfil en ORCID, esto se vea reflejado automáticamente en el sistema y en el repositorio y viceversa. La producción científica estaría almacenada únicamente en NINIVE convirtiéndose así en una fuente institucional de información. La hoja de vida del investigador se mantendrá actualizada en el SIGI y visible a través de ORCID, redes sociales, buscadores y agregadores especializados. Al contar con una hoja de vida integrada y actualizada, el investigador podrá generar diferentes currículums a través de servicios web y/o herramientas de exportación de datos, para facilitar el registro y actualización de la información requerida para los Programas de Beca al Desempeño Docente e Ingreso, Permanencia y Promoción de la Institución y para las convocatorias PRODEP y SNI. En la Figura 2 se muestra el modelo y arquitectura del sistema.



**Figura 2. Modelo y arquitectura del Sistema Institucional de Gestión de la Investigación de la U.A.S.L.P.**

**Fuente: Elaboración propia**



## Avances y resultados del proyecto piloto

El modelo del sistema SIGI fue diseñado y presentado en el segundo semestre del 2015 tanto a la Secretaría de Investigación como a la Secretaría Académica, responsables respectivamente de la gestión de la investigación y administración de la docencia de la UASLP. A principios del 2016, se presentó el proyecto a una comunidad de investigadores contando con la participación de la representante de ORCID de América Latina. A partir de ahí, se conformó un grupo piloto de 20 investigadores de diferentes dependencias académicas que participarán en la retroalimentación y validación de la prueba piloto.

Para la fundamentación del proyecto, una de las primeras acciones llevadas a cabo para identificar las buenas prácticas en el desarrollo de sistemas CRIS y su interoperabilidad con un RI, fue el análisis de diferentes estudios de caso presentados en los eventos internacionales de CRIS2016<sup>10</sup> y Open Repositories 2016<sup>11</sup>.

Adicionalmente, en el mes de junio se visitaron dos universidades españolas, la Universidad de Barcelona (UB) y la Universidad Politécnica de Catalunya (UPC). En el caso de la UB, su sistema CRIS llamado GREC (<http://webgrec.ub.edu>) está conectado al Repositorio Institucional de la UB mediante el registro de los diferentes IDs de sus investigadores (ORCID, Research ID y Scopus ID), la creación de un catálogo de las revistas en donde publican sus investigadores y la implementación de APIs y servicios web para la recuperación de información de bases de datos como Scopus y Web of Science. Por su parte, la UPC cuenta con un conjunto de sistemas y servicios interconectados, el Repositorio UPCommons (<http://upcommons.upc.edu/>), el CRIS denominado DRAC (<http://drac.upc.edu/>), el portal PUBLICA! (<http://publica.upc.edu/>) y el Portal FUTUR de la Producción Científica de los Investigadores de la UPC (<http://futur.upc.edu/>).

Por otro lado, respecto a la implementación del prototipo del sistema, se han hecho pruebas con la API pública ORCID, considerando la posibilidad de contratar la membresía institucional para obtener un mayor soporte y capacidad de creación y gestión de identificadores para la comunidad de investigadores de la UASLP. También se han efectuado y documentado pruebas con las APIs y plataformas de Scopus, Web of Science, Crossref y la actualización por medio de archivos BiTex. Hasta el momento (agosto 2016) se ha diseñado una primera interfaz que contiene los datos generales del investigador y los registros con las referencias a los artículos a partir del perfil público de ORCID y la información de las bases de datos institucionales. Además, se está evaluando el software DSpace-CRIS, como una posible solución para implementar total o parcialmente el sistema SIGI con la estructura actual del Repositorio NINIVE y su interoperabilidad con ORCID.

## Conclusiones

Aunque se cuenta con la información, experiencia y soporte necesario para la implementación de la solución tecnológica, el éxito de la prueba piloto dependerá en gran medida del grado de involucramiento del grupo de investigadores que mediante su colaboración y retroalimentación hagan posible el flujo de trabajo propuesto. Una vez que se concluya la prueba piloto, los resultados globales del proyecto serán presentados a la Rectoría de la UASLP para su aprobación y eventual implementación.

<sup>10</sup> CRIS2016.

[https://www.eventsforce.net/STANDREWS/media/uploaded/EVSTANDREWS/event\\_43/CRIS2016\\_Final\\_Programme\\_With\\_Poster\\_List\\_Session\\_Chairs\\_Hyperlinks.pdf](https://www.eventsforce.net/STANDREWS/media/uploaded/EVSTANDREWS/event_43/CRIS2016_Final_Programme_With_Poster_List_Session_Chairs_Hyperlinks.pdf)

<sup>11</sup> Open Repositories 2016. <https://www.conftool.com/or2016/sessions.php>

## Referencias

Castro, P., Shearer, K. & Summann, F. (2014). The gradual merging of repository and C R I S solutions to meet institutional research information management requirements. CRIS 2014. Science Direct. Recuperado el 15 de noviembre de 2015 desde: [http://dspacecris.eurocris.org/bitstream/11366/197/1/DeCastro\\_Shearer\\_Summann\\_CRIS2014\\_Rome.pdf](http://dspacecris.eurocris.org/bitstream/11366/197/1/DeCastro_Shearer_Summann_CRIS2014_Rome.pdf)

DURASPACE. Recuperado el 30 de junio de 2016 desde: <https://wiki.duraspace.org/display/DS-PACECRIS/Dspace-CRIS+Home>

EuroCRIS(2016). Recuperado el 15 de junio de 2016 desde: <http://dspacecris.eurocris.org/>

Swan, A. (2008). The business of digital repositories. Recuperado el 8 de diciembre de 2013 de <http://eprints.ecs.soton.ac.uk/14455>

# Visibilidad

Malgorzata Lisowska (URosario)

# “Detección de bots en reportes estadísticos”

*Catá, Juan Manuel ; Lira, Ariel Jorge ; De Giusti, Marisa Raquel*

---

## Resumen

Las estadísticas de un repositorio institucional son una herramienta básica que asiste el proceso de toma de decisiones y gestión del repositorio. Por este motivo, es importante que la información provista por estas estadísticas sea información precisa y confiable, en particular los registros de acceso y descarga.

Los repositorios digitales concentran una gran cantidad de enlaces entrantes y muchos contenidos de calidad por lo que resultan de mucho interés para los bots que navegan la World Wide Web. Si bien la mayoría de los bots respetan las reglas básicas establecidas en los archivos robots.txt, muchos de ellos no lo hacen e incluso hay algunos que no se identifican como tales y se hacen pasar por agentes de usuario normales. A pesar de las medidas que se toman para evitar el acceso de bots maliciosos, un número importante de estos logra filtrarse y efectuar miles de accesos indeseados. Se genera, en consecuencia, gran cantidad de datos espurios que llevan a estadísticas poco fiables y que en última instancia entorpecen el proceso de gestión del repositorio

Para solucionar el problema planteado, se comenzó desarrollar una mecanismo que, a partir del análisis, permita filtrar los accesos de bots normales y bloquear los accesos de bots maliciosos o con mal comportamiento.

Las pruebas iniciales con la herramienta han permitido identificar un número elevado de accesos correspondientes a bots maliciosos que, al filtrarlos, permiten obtener resultados estadísticos mucho más veraces.

## Abstract

Statistics are an essential tool for institutional repositories which assists the decision making process and repository management. Therefore, the information they provide must be precise and reliable, specially those based on access and download logs.

When a digital repository grows and brings together large amounts of incoming links and high quality content, it acquires of great significance for bots. Most bots follow the basic rules established in robots.txt files, nevertheless many do not do it, and some of them do not identify themselves as bot masquerading as normal users. Despite the measures taken to avoid access to malicious bots, a large amount of them manage to seep and make thousands of unwanted access. Therefore a large number of spurious data is generated which leads to unreliable statistics and hinders the repository management process.

In order to solve this problem, a mechanism was developed to analyse, detect and filter access from malicious or misbehave bots.

Initial tests with this tool allowed to identify a large number of access comming from malicious bots that, after being filtered, allows to get much more sound and reliable statistics results.

## Introducción

Las estadísticas de un repositorio institucional son una herramienta indispensable para la obtención de métricas que definen el alcance y el impacto del repositorio y de cada recurso dentro del mismo. Algunos reportes típicos podrían ser:

- Cantidad de accesos a un recurso o colección
- Documentos más descargados
- Accesos de los recursos de una colección
- Distribución de accesos por origen (continente/país/ciudad)
- Tasa de accesos por fecha
- Fechas con mayor cantidad de accesos
- entre otros

Estas métricas influyen la gestión y la toma de decisiones en los repositorios, por tal motivo es de suma importancia que los datos obtenidos sean fiables y precisos.

DSpace incluye un módulo estadísticas (Estadísticas DSpace, 2016) que ofrece reportes simples a nivel global, de comunidad, de colección y de ítem de:

- ranking de cantidad de accesos por país y ciudad
- cantidad total de accesos/descargas
- ranking de ítems más accedidos
- 

El módulo está basado en el software de indexación de texto Apache Solr que registra todos los eventos vinculados al acceso al repositorio, a la actividad interna de su flujo de trabajo y al uso del módulo de búsqueda. Los datos más importantes que se asocian con los eventos de acceso y que se utilizan para el presente trabajo son:

- type: tipo de recurso accedido (ítem, comunidad, colección, etc)
- id: combinado con el type identifica el recurso.
- ip: IP desde donde proviene el acceso
- uid: id interno del objeto de DSpace (a partir de DSpace 6)
- continent: continente desde el cual se accede
- country: país desde el cual se accede
- city: ciudad desde donde se accede
- userAgent: agente de usuario utilizado al momento de acceder
- isBot: identifica si la ip pertenece a un bot o no
- referrer: identifica la dirección de la página que creó el vínculo con el recurso que está siendo solicitado
- además de muchos otros campos complementarios que aún no se utilizan como parte del análisis.

Este módulo registra todos los accesos, tanto los provenientes de bots como de usuarios normales, y utiliza una etiqueta isBot para diferenciar los accesos de cada caso. El etiquetado de los accesos de bots se realiza de dos formas: 1) a partir de una lista de direcciones IP de bots conocidas y 2) a partir del header HTTP *userAgent* que proviene en la solicitud.

Sin embargo, a pesar de disponer de estas formas de detectar y filtrar el acceso de bots, diariamente acceden al sistema bots malintencionados (spiders, miners, crawlers,

etc), que no se identifican como tales, y buscan acceder y descargar todo el contenido posible del repositorio, generando un número muy alto de accesos y contaminando los registros estadísticos.

Entre otros, se identifican 4 grandes grupos de bots que fueron tenidos en cuenta durante el desarrollo del prototipo:

1. Bots normales conocidos, como el de Google (GoogleBot), que respeta las reglas especificadas en el robots.txt. El módulo de estadísticas de DSpace ya incluye la funcionalidad para detectar y descartar estos casos a partir del procedimiento mencionado.
2. Bots normales aunque desconocidos: son bots que se identifican como tales, por ejemplo, mediante el user agent, y respetan el archivo robots.txt pero que no son conocidos en el sistema porque son nuevos o muy específicos.
3. Bots con comportamiento incorrecto: son procesos que buscan acceder y descargar todo o parte del contenido del repositorio sin respetar las reglas básicas establecidas en el archivo robots.txt. Este tipo de bots son muy perjudiciales porque como suelen realizar numerosos accesos concurrentes, suelen degradar la performance del sistema y hasta pueden provocar su caída.
4. Bots maliciosos: típicamente spiders o miners que buscan vulnerar el sistema para conseguir información protegida, publicar datos no autorizados, publicidad, o simplemente afectar negativamente el repositorio.

## Propuesta

A partir de la problemática planteada y de la funcionalidad provista por Apache Solr y DSpace, se inició el desarrollo de una herramienta modular y configurable, integrada al módulo de manejo de estadísticas de DSpace (stats-util) que, mediante el análisis de los eventos de accesos registrados en el módulo de estadísticas, intenta:

1. detectar direcciones IP con comportamientos anómalos característicos de bots no identificados y marcar los eventos asociados con dicha dirección con el flag isBot
2. Guardar la dirección detectada en una lista de direcciones IP de bots conocidos para que sean detectados de forma temprana en futuros accesos.(AUTOR, 2008)

Un esquema de trabajo similar se plantea en (Stassopoulo A. y Dikaiakos M. D 2006)” en el cual se propone un mecanismo para evaluar los registros de accesos o logs provistos por servidores web para detectar la actividad de *crawlers*.

El proceso de detección ejecuta una serie de Reglas, cada una de las cuales busca detectar patrones de comportamiento anómalos específicos y, de acuerdo con la probabilidad calculada de que las direcciones IP detectadas sean bots, marcarlas automáticamente o bien generar un reporte que pueda ser analizado posteriormente por un humano para definir si realmente se trata de un bot, o si es un caso aislado y particular. La funcionalidad de la herramienta no es filtrar ni bloquear el acceso de bots al sistema, sino mantener los registros estadísticos sanos e identificar cuánto de lo que se registró corresponde a accesos de bots, para que las próximas métricas que se realicen basadas en estas estadísticas, sean lo más correctas posibles.

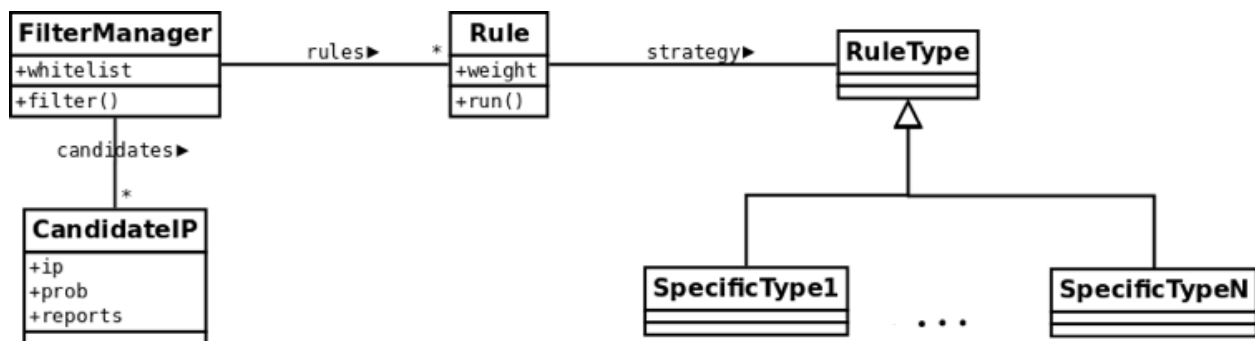


Este trabajo se está llevando a cabo sobre CIC-Digital, el repositorio institucional de la Comisión de Investigaciones Científicas (CIC) de la provincia de Buenos Aires (Argentina), y se prevé su aplicación luego en el repositorio institucional de la UNLP, SEDICI.

## Implementación

La herramienta consta de un controlador central, encargado de ejecutar las reglas especificadas en un archivo de configuración, y mantiene una lista con las direcciones IP candidatas junto con la probabilidad de que dicha dirección IP sea un bot. Una vez ejecutadas todas las reglas, el controlador informa los resultados y marca las direcciones IP detectadas como bots en los registros estadísticos.

El modelo es ligeramente más complejo, pero podría resumirse de la siguiente manera:



Las reglas son elementos configurables encargados de analizar los eventos de acceso del módulo de estadísticas, para evaluar la actividad de cada dirección IP y en caso de detectar una dirección IP sospechosa, se le asigna a esta una “probabilidad de bot” de acuerdo a cuán seguro se esté de que dicha IP sea un bot. Por ejemplo, si se detecta que una dirección IP tuvo 100 accesos en un mes, la probabilidad asociada es baja, en cambio si tuvo más de 1000 accesos, la probabilidad es mucho mayor (los valores usados son a modo de ejemplo, cada repositorio debería manejar los valores correspondientes a sus propias necesidades).

## Tipos de reglas

El tipo de regla (*RuleType*) contiene la implementación de la lógica de búsqueda que tendrá una regla. La herramienta permite definir varias reglas con un mismo *RuleType* a las cuales se le especifican parámetros diferentes que se adecúen al caso buscado, por ejemplo: un *RuleType* que busca cantidad de accesos<sup>1</sup> en un período de tiempo, podría ejecutarse una vez para analizar los accesos a ítems, y otra para la cantidad de descargas ese período de tiempo.

Una de las ventajas más importantes de la utilización de estas reglas como módulos individuales es la adaptabilidad y la extensibilidad. A futuro podrían crearse nuevas estrategias de búsqueda que se adapten mejor al repositorio que se está analizando. Uno de los tipos de reglas creados hasta el momento se encarga de evaluar la cantidad de descargas (o accesos) en períodos fijos de tiempo. La regla es completamente configurable y permite ajustar el lapso de tiempo, la cantidad de accesos mínimos requeridos para considerar una dirección IP sospechosa, el tipo de acceso (bitstream, item, collec-

<sup>1</sup> Los accesos se diferencian por tipo, (item, bundle, collection, etc). Cuando se habla de cantidad de accesos podría ser cualquiera de estos tipos

tion, etc), entre otros ajustes. Este tipo de regla permite detectar secuencias de acceso que sería imposible de realizar para un humano.

Con este criterio se detectaron por ejemplo muchos casos de bots no identificados que realizan:

- cientos de descargas por hora y absorben todos los archivos PDF del sitio
- miles de accesos continuos períodos largos como un mes

También se definió otro tipo de regla que realiza un análisis por subred a partir de la identificación de accesos repetidos desde direcciones de IP con raíz similar. Si se detectan muchos accesos provenientes de una misma subred podría tratarse de una botnet y debe ser tenido en cuenta.

## Configuración

La definición de la heurística de diagnóstico es un proceso extremadamente dinámico dado que hay varios factores que requieren ajustar, agregar o eliminar reglas de forma permanente. Algunos factores son:

- se debe realizar un ajuste de los parámetros de cada regla
- se pretende usar el desarrollo en varios repositorios y no todos los repositorios reciben los mismos bots
- el comportamiento de los bots puede ser cambiante y esporádico
- la ejecución de algunas reglas muy complejas pueden afectar la performance del sistema

Por tal motivo, es necesario que la herramienta en general y las reglas en particular se adapten para cubrir las necesidades particulares de cada caso. A continuación se detallan algunas configuraciones posibles:

- Las cadena de reglas que serán ejecutadas en orden
- La configuración específica de cada regla: se indica el peso de la misma (valor que determina la probabilidad de bot) y parámetros complementarios que dependen directamente del tipo de regla a ejecutar. Por ejemplo, una regla de cantidad de accesos por mes tendrá los siguientes parámetros propios definidos:
  - » Mínimo de accesos para considerar la dirección de IP como sospechosa
  - » Tipo de recurso a considerar
  - » Período de tiempo en el cual se evalúa la condición
- Una lista segura o whitelist de direcciones IP que no deben ser tenidas en cuenta al momento de analizar los registros porque es garantizado que son confiables y no son bots. Pueden existir casos de direcciones IP con un comportamiento inválido para un usuario normal, pero que no obstante eso de antemano se conoce que no son bot como por ejemplo: la subred institucional desde la que los administradores ingresan cientos de veces por día, una ip de una red NAT, entre otros.

## Caso de prueba

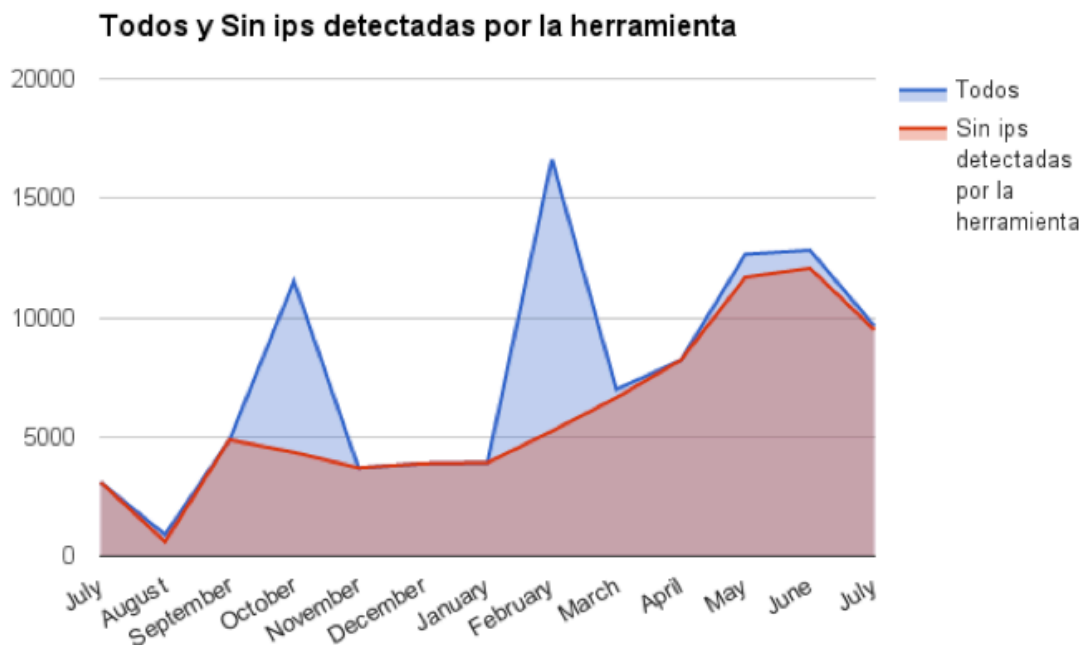
En primer lugar se ejecutó la herramienta sobre el repositorio CIC-Digital analizando la cantidad de accesos a ítems por hora que se realizaron entre julio de 2015 y julio del 2016, tomando como criterios de búsqueda aquellas direcciones IP que accedieron a más de 50 ítems por 1 hora. Se detectaron direcciones IP que efectuaron desde 60 accesos en una hora, otras con más de 300, e incluso un caso con más de 1000 accesos en 1 hora.

También, como se menciona en (Balla et Al, 2011) la hora a la que se realizan dichos accesos puede ser un factor importante, en el artículo se habla de night requests y se refiere a los accesos realizados entre las 2am y las 8am.

En segundo lugar se realizaron consultas sobre los registros de acceso del módulo de estadísticas para generar un gráfico filtrando según 2 criterios, a saber:

1. Todos los accesos realizados entre el 01/07/2015 y 01/07/2016.
2. Todos los accesos en el mismo período sin considerar los accesos de las IP sospechosas detectadas previamente con la herramienta.

Como se muestra en el gráfico a continuación, cuando se consulta por todos los accesos realizados al sistema se ven picos de accesos sospechosamente altos en los meses de octubre del 2015 y febrero del 2016 teniendo en cuenta los accesos que se venían registrando. Luego de omitir las direcciones IP detectadas por la herramienta, los resultados obtenidos cambiaron drásticamente, se eliminaron los picos de octubre y febrero, y se aprecia una leve mejora en los meses de agosto del 2015, marzo, mayo y junio del 2016.



## Conclusiones

Las pruebas preliminares de la herramienta permiten observar numerosos casos de direcciones IP pertenecientes a bots no identificados que acceden a items y descargan contenido del repositorio, los cuales registran números muy elevados de accesos, contaminando seriamente los registros estadísticos. Luego de filtrar dichas direcciones IP se logró conseguir estadísticas mucho más fiables.

Los resultados preliminares convalidan el sentido del desarrollo y obligan a continuar con el refinamiento de la herramienta, particularmente para reducir la cantidad de falsos negativos que aún escapan al detector y hacer un análisis más exhaustivo de los resultados obtenidos para prevenir falsos positivos.

La flexibilidad de configuración y el beneficio en la calidad de estadísticas resultantes hacen de éste un producto muy útil para cualquier repositorio basado en DSpace. Por tal motivo, se prevé proponer el desarrollo para que sea integrado a DSpace en un futuro cercano.

## Bibliografía

Stassopoulo A. y Dikaiakos M. D. (2006) Crawler Detection: A Bayesian Approach. Proc. Int'l Conference on Internet Surveillance and Protection (ICISP'06), 16-21.

doi:10.1109/ICISP.2006.7.

<http://linc.ucy.ac.cy/publications/pdfs/2006-CISP-CrawlerDetection.pdf>

Balla, A.; Stassopoulou, A.; Dikaiakos, M. D. (2011) Real-time Web crawler detection. Proc. of 18th International Conference on Telecommunications (ICT'11), pp. 428-432. doi:10.1109/CTS.2011.5898963. <http://linc.ucy.ac.cy/publications/pdfs/2011-ICT-ANapa.pdf>

Modulo de Estadísticas DSpace - <https://wiki.duraspace.org/display/DSDOC5x/SOLR+Statistics>  
Revisado el 8 de agosto de 2016.

CIC-digital - <http://digital.cic.gba.gob.ar> (revisado al 9/8/2016)

# “Repositórios institucionales en Brasil”

*Josimara Dias Brumatti. Simone da Rocha Weitzel.*

---

## Resumo

A pesquisa tem por objeto de estudo a produção científica de Instituição de Ensino Superior publicada em periódico e depositada em Repositório Institucional, visando caracterizar o Acesso Aberto Verde no Brasil. A produção científica será classificada em vertentes elaboradas a partir da Declaração de Budapeste (2002) e Harnad (2004; 2012) para mapear o tipo de acesso dos artigos, predominantes das áreas do conhecimento estudadas, inseridos nos repositórios nacionais. Pressupõe que o desempenho entre áreas do conhecimento no Acesso Aberto, a avaliação de desempenho da Pós-Graduação brasileira, assim como as particularidades na produção científica no Brasil, são impasses que afetam o Acesso Aberto Verde<sup>1</sup> (AA Verde). Para o alcance deste propósito, baseia-se em Ortellato (2008), Furnival, Oprime e Silva Perez (2013), Alperin, Fischman e Willinsky (2008), Terra-Figari (2008), Guédon (2010) e Vessuri (2015).

**Palavras-Chave: Acesso Aberto; Produção Científica; Repositório Institucional; Brasil.**

## Abstract:

The research aims to scientific production institution of higher education published in journal and placed in the Institutional Repository, to characterize the Open Access Green in Brazil. The scientific production will be classified in sections prepared from the Budapest Open Access Initiative (2002) and Harnad (2004; 2012) to map the access of articles, predominant areas of knowledge studied, embedded in national repositories. Assumes the performance between areas of knowledge in Open Access, the assessment of Brazilian Graduate performance, as well as the particularities of scientific production in Brazil are impasses affecting AA Green. To achieve this purpose, is based on Ortellato (2008), Furnival, Oprime and Silva Perez (2013), Alperin, Fischman and Willinsky (2008), Terra-Figari (2008), Guédon (2010) and Vessuri (2015).

**Keywords: Open Access; Scientific Production; Institutional Repository; Brazil.**

## Introdução:

O AA é o meio de democratizar a produção científica publicada essencialmente em periódico e financiada com recursos públicos. Duas estratégias foram definidas na Budapest Open Access Initiative (BOAI, 2002): 1. Periódicos de Acesso Aberto e 2. Auto-arquivamento da produção científica em Repositórios, sendo intituladas por Harnad e outros (2004) de Via Dourada e Via Verde, respectivamente.

<sup>1</sup> Estratégia do AA chamada anteriormente à 2012 de Via Verde.

A Via Dourada estabelece o acesso por meio de Periódicos de Acesso Aberto, considerando os periódicos que são livres de taxas, tanto para publicação pelo autor, e/ou taxas de acesso para o leitor. A Via Verde é baseada no depósito da produção científica publicada em periódico comercial, especialmente artigos de periódicos.

A Via Verde é apontada pelos signatários e ativistas do Movimento do Acesso Aberto no cenário mundial, como Harnad (2004, 2012, 2015), Guedón (2010), Suber (2004, 2015), Rentier (2010), Rodrigues (2004, 2014, 2015), como estratégia central para o desenvolvimento do AA.

No Brasil, Leite (2009, 2011), Kuramoto (2004, 2014), Weitzel (2006, 2014), também indicam a Via Verde como o principal caminho para dar acesso ao conhecimento produzido no âmbito dos institutos de pesquisa e universidades, especialmente as públicas. A produção científica nacional é atrelada em sua maioria a Instituições de Ensino Superior (IES) pública que recebem financiamento governamental para realização de pesquisas, por tanto é justo que os resultados das pesquisas oriundos destes financiamentos estejam acessíveis para toda a sociedade.

Este estudo parte do pressuposto que a produção científica brasileira se difere da produção da ciência mundial por possuir características específicas e singulares que envolvem o modelo de CC brasileira por fazer parte da Ciência Periférica, podendo acarretar uma distinção no desenvolvimento da Via Verde e Via Dourada.

Para tanto, é oportuno realizar um estudo que analise a produção científica brasileira para mapear o atual cenário do desenvolvimento da Via Verde no Brasil, ratificando sua importância como instrumento de fomento à disseminação científica, democratização da informação pública e preservação da memória institucional no contexto do AA.

Esta pesquisa tem por objetivo caracterizar a Via Verde no Brasil a luz do conceito de AA definido pela BOAI (2002) e Harnad e outros (2004) por meio de estudo exploratório mapeando a produção científica oriunda de Instituição de Ensino Superior (IES) brasileira publicada em periódico e o respectivo arquivamento dessa produção em Repositórios Institucionais (RIs).

Apontamos para os seguintes aspectos como justificava para a presente pesquisa, baseando-se nos seguintes autores: A) Desempenho distinto na Via Verde no Brasil nas Áreas do Conhecimento: Ortellado (2008); Furnival; Oprime e Silva Perez (2013). B) Critérios da avaliação de desempenho da Pós-Graduação no Brasil: Ortellado (2008); Furnival (2013) e C) Modo de comunicar a ciência: Alperin, Fischman e Wilinsky (2008); Terra-Figari (2008); Guedón (2010); Vessuri (2015).

Ortellado (2008) demonstra que apesar do sucesso no AA no Brasil, houve um déficit no desempenho em determinadas áreas do conhecimento. Segundo o autor, as políticas nacionais de AA “são ainda pouco eficazes, tendo um impacto pequeno nas ciências naturais em face do alto grau de internacionalização das publicações”. (ORTELLADO, 2008).

Furnival, Oprime e Silva Perez (2013), examinam a percepção do AA entre pesqui-



sadores brasileiros e fatores que afetam a aceitação ou resistência ao AA pelos autores. Em relação à obrigatoriedade dos pesquisadores disponibilizarem seus trabalhos em AA através de políticas mandatárias, o estudo demonstra que aproximadamente 35% são favoráveis ao depósito, no entanto 53 % são contrários de alguma forma à imposição do depósito.

Quando a análise de Furnival, Oprime e Silva Perez (2013) mapeia a obrigatoriedade de depósito através mandatos de AA, há concordância desigual entre áreas do conhecimento. Enquanto que mais de 50% das áreas de Ciências Humanas e Sociais concordam com a obrigatoriedade do depósito, apenas 17% da Ciência Exata é favorável.

Para os mesmos autores, os pesquisadores brasileiros têm preferência por publicarem em periódicos com alto fator de impacto, pois além de serem mais prestigiadas são condições para obter financiamentos e verbas de órgãos públicos de fomento, dando enfoque na avaliação somente a dados quantitativos.

Segundo Terra-Figari (2008), os países da “periferia científica” em especial na América Latina, são prejudicados por desvantagens em: tamanho; incentivo; financiamento; linguagem de publicação e edição de parâmetros de publicação internacional.

Para Guédon (2010) o AA corrige muitas das injustiças inerentes à barreira de conhecimentos existentes entre os chamados países centrais e periféricos, tornando-se uma possibilidade inegável.

Para Vessuri (2015) o que determina a baixa participação da América Latina na produção mundial não é somente a ciência não ser *mainstream*. A ausência de revistas científicas latino americanas nas principais bases de dados de indexação e a barreira da língua interferem em seu progresso científico. Esses fatores, assim como o baixo investimento governamental em inovação e desenvolvimento, impedem que haja uma competição igualitária entre os países ditos periféricos.

Segundo Alperin, Fischman e Willinsky (2008) as motivações da América Latina para promoção do Acesso Aberto à produção científica são diferentes das motivações dos países centrais. Enquanto na Europa, Estados Unidos e Canadá, o AA tem sido impulsionado em parte, por um aumento nos preços e na quantidade de periódicos, as publicações na América Latina têm sido tradicionalmente livres.

Para Guédon (2010) o AA abre perspectivas plenas de esperanças, “isto é, com o Acesso Aberto, corrigir muitas injustiças inerentes à barreira de conhecimentos existente entre os chamados países centrais e periféricos torna-se uma possibilidade inegável”. (GUÉDON, 2010, p. 73).

Pretende-se com este estudo caracterizar a via verde no Brasil e mapear possíveis obstáculos que limitam a via a fim de contribuir para o fomento de ações para chegar ao 100% AA. Além disso, permitirá apontar possíveis meios para que o Brasil desenvolva iniciativas bem-sucedidas na Via Verde.

## Acesso Aberto Verde no Brasil

No Brasil as ações para tornar acessíveis as pesquisas publicadas em periódicos iniciaram antes do MAA, através de estratégias como o SciELO, que surgiu em 1997. No entanto com o encadeamento das iniciativas em AA pelo mundo, o Brasil se inseriu no MAA, mais precisamente pela participação de Hélio Kuramoto, Gestor do IBICT na época e Sely Costa da Universidade de Brasília, na 7ª *International Conference on Electronic Publishing* – ELPUB, “aos quais, ao retornarem ao Brasil, iniciaram conversações e gestões visando à inserção do Brasil no movimento” (KURAMATO; COSTA; LEITE, 2010, p. 135).

O IBICT desenvolveu ações em parceria com outros órgãos de financiamento e viabilizou recursos para projetos com iniciativas de consolidação do AA no Brasil. Dentre os projetos de promoção do AA podemos indicar dois principais de cunho tecnológico que abordaram o AA Verde e Dourado: a) customização do sistema *Open Journal Systems* – OJS que originou o Sistema Eletrônico de Editoração de Revistas – SEER em 2003, para gestão de periódicos científicos *on-line*; b) o projeto IBICT-FINEP/PCAL/XBDB, que possibilitou a implantação de aproximadamente 40 repositórios institucionais. O projeto lançado em 2009 consiste na distribuição de kits tecnológicos, no treinamento dos recursos humanos da instituição e no suporte informacional e técnico para o bom desenvolvimento destes sistemas. (KURAMATO; COSTA; LEITE, 2010) (IBICT, 2016).

Apesar de iniciativas nacionais para o desenvolvimento das duas estratégias propostas pelo MAA, a consolidação do AA vem ocorrendo de formas diferentes em países desenvolvidos, principalmente na América Latina. Segundo Costa e Leite (2016) as iniciativas em Acesso Aberto Dourado (AA Dourado)<sup>2</sup> avançaram significativamente na região, “esta constatação vem de iniciativas bem-sucedidas tais como SciELO e adoção generalizada do OJS”.

Ainda segundo Costa e Leite (2016), apesar dos avanços recentes do AA Verde, em particular o aumento do número de RIs e a intensificação do trabalho de cooperação na região, ainda há muito a ser feito nessa área.

Em 2011 o Senador da República Rodrigo Rollemberg submete ao Senado Nacional o Projeto de Lei - PLS 387/2011<sup>3</sup>, que dispõe sobre o processo de registro e disseminação da produção técnico-científica pelas instituições de educação superior, bem como as unidades de pesquisa no Brasil. Cinco anos depois o PLS ainda está em tramitação, pronta para a Pauta na Comissão, no entanto em Fevereiro de 2015 o Senador Álvaro Dias, votou pela rejeição do projeto, alegando injuridicidade e inconstitucionalidade.

O contexto nacional fez com que Universidades e Organizações Públicas de pesquisa brasileiras transpassassem a delonga da tramitação política e adotaram mandatos institucionais de AA. A Fundação Oswaldo Cruz - FIOCRUZ aprovou uma política mandatária a nível institucional, no entanto Rodrigo Murinho, vice-diretor de comunicação do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde - ICICT, deixa clara a importância de uma política nacional.

<sup>2</sup>Estratégia do AA chamada anteriormente à 2012 de Via Dourada.

<sup>3</sup>Acesso o texto na íntegra: <http://www.senado.gov.br/atividade/materia/getPDF.asp?t=93151&tp=1>.

Uma política tende a fortalecer as iniciativas institucionais, uma vez que exige a disponibilização dos resultados de estudos e pesquisas para a sociedade, principalmente quando os recursos para a realização desses trabalhos acadêmicos são públicos. Esse é o caso das leis da Argentina e do Peru, de 2013; e do México, de 2014. Em todos os países foram instituídos repositórios nacionais, congregando universidades e instituições de pesquisa. (MURTINHO, 2015).

Carlos Machado de Freitas, coordenador do Portal de Periódicos da Fiocruz, destaca que quase todas as pesquisas no Brasil são realizadas em instituições públicas, financiadas principalmente com recursos públicos e que, por isso, são bens públicos.

No caso do Brasil, uma política nacional de acesso aberto não só é coerente com a estrutura e o financiamento das pesquisas, como também implica em reforçar nosso compromisso com a democratização de nossa produção e acesso ao conhecimento. (FREITAS, 2015).

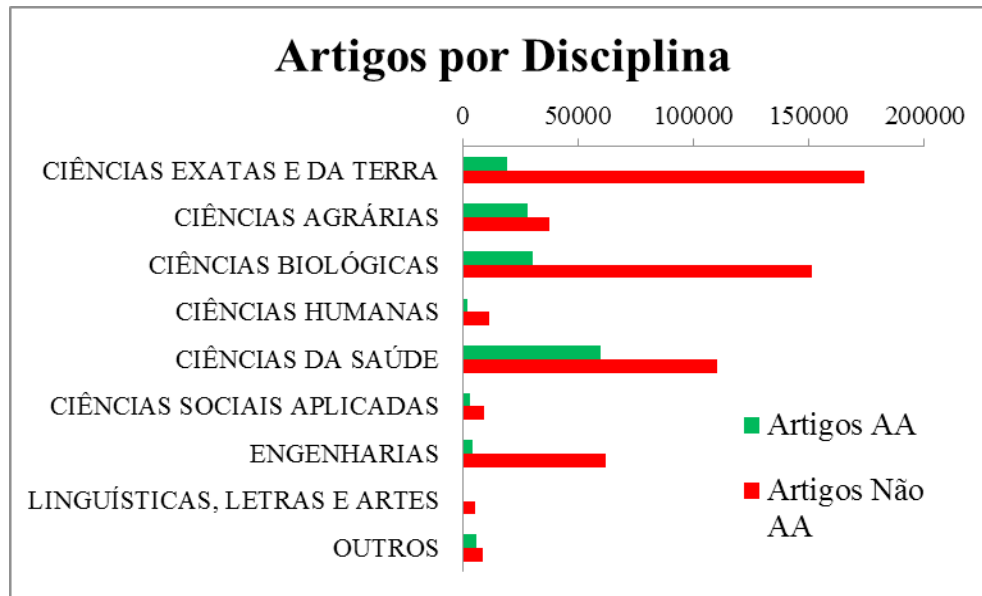
Essa visão pública e governamental do AA tem encontrado barreiras em diversos países no mundo, pois penetram questões políticas e jurídicas. Contudo no Brasil, ainda caminhamos com iniciativas isoladas e pouco abrangentes.

Ortellado (2008) menciona como fator de desigualdade no desenvolvimento da Via Verde no Brasil, a produção científica de algumas áreas do conhecimento, como as Ciências Naturais e da Terra e as Ciências Humanas e Sociais.

Enquanto as Ciências Sociais se publica mais em periódicos brasileiros, nas Ciências Naturais se publica mais em periódicos estrangeiros.... Trata-se também de um problema cada vez maior para o Brasil e outros países periféricos à medida que a sua produção científica (em especial nas Ciências Naturais) melhora de qualidade e busca a exposição nos periódicos indexados ou bem avaliados pelas agências financiadoras brasileiras (CAPES). (ORTELLADO, 2008).

Os dados abaixo foram extraídos da base de dados comercial Web of Science. O gráfico 2 mostra os artigos indexados na plataforma dividido por área do conhecimento. Através destes dados, podemos visualizar o desenvolvimento do AA por área do conhecimento no mundo.

Gráfico 1: Artigos por área do conhecimento<sup>4</sup> indexados na Web of Science no mundo



Fonte: O autor (2016).

Ainda segundo Ortellado (2008), a diferenciação entre as áreas do conhecimento acima mencionadas gera impacto em iniciativas no AA.

O efeito desse desigual padrão é que as políticas nacionais de estímulo às publicações de AA têm impacto diferente nas diversas áreas do conhecimento. Enquanto nas Ciências Sociais essa política tem conseguido liberar o acesso a quase metade dos artigos, nas Ciências Naturais os resultados são mais modestos. (ORTELLADO, 2008).

Sendo a Via Dourada reconhecida no cenário brasileiro por iniciativas de destaque como o SciELO, segundo Ortellado (2008) “o país descuidou das políticas” governamentais para o desenvolvimento e maior adesão da Via Verde. Portanto embora o país lidere o ranking<sup>5</sup> mundial de artigos periódicos científicos em AA, onde a publicação “nasce” em AA, existem ainda contextos em que a publicação tem o acesso restrito sendo fundamental a Via Verde para torná-la disponível amplamente.

Adicionado ao desequilíbrio no desempenho do AA nas áreas do conhecimento, podemos indicar que a avaliação de desempenho pode influenciar o AA, mais precisamente a avaliação de desempenho da pós-graduação, que tem seu foco na análise quantitativa de artigos de periódicos.

A avaliação da Pós-Graduação no Brasil é realizada pela CAPES, que tem por competência “coordenar a política do sistema nacional de pós-graduação e exercer a função de incentivar, em nível de mestrado e de doutorado, a expansão e consolidação dos programas de pós-graduação”. (CASTANHA; GRÁCIO, 2012).

Visa, ainda, assegurar a validade nacional desses diplomas e o avanço dos cursos, por meio da Avaliação Trienal dos Programas de Pós-graduação do país, que utiliza importantes indicadores destinados a acompanhar o desen-

4 As áreas do conhecimento foram elencadas segundo a tabela do conhecimento do CAPES.

5 O Brasil está em 2º no ranking DOAJ em número de títulos de revistas indexados, atrás apenas dos Estados Unidos.

volvi-mento, evolução e consolidação dos programas, contribuindo para a formação de recursos humanos com alta competência científica e para a promoção da co-laboração científica internacional. (CASTANHA; GRÁCIO, p. 84, 2012).

São considerados 5 fatores como critérios para avaliação da pós-graduação, a saber: 1) Proposta do programa; 2) Corpo docente; 3) Corpo discente, teses e dissertações; 4) Produção intelectual; 5) Inserção social, sendo a pontuação variada segundo as áreas do conhecimento. Para Voguel e Kobashi (2015), há um sexto critério na avaliação, que seria a Internacionalização, que não fica explícito, mas que é apresentado nos Documentos de Área como item de distinção entre os programas de excelência.

O critério de avaliação da “Proposta do Programa” tem peso zero. O “Corpo Do-cente” gira entre 15 a 20% da nota, o “Corpo Discente, Teses e Dissertações” correspondem de 30 a 35% do total, a “Produção Intelectual” representa 35 a 40% da avaliação e a “Inserção Social” equivale a 10 a 15% da nota. O quesito de Internacionalização aplica-se “somente a programas de doutorado que já atingiram um excelente nível nos demais quesitos, recebendo nota cinco, mas que ainda podem ser diferenciados por suas ações com outros países. Assim, este quesito destina-se a atribuir a nota 6 ou 7 a um programa”. (VOGUEL E KOBASHI, 2015).

Ainda há o critério de avaliação o Qualis Periódicos, que incide na estratificação da “Produção Intelectual” que é feita de acordo com uma escala de pontuação do periódico no qual os artigos foram publicados. A qualidade do periódico, por sua vez, pode ser conferida de acordo com a base de dados na qual o periódico foi indexado. (VOGUEL E KOBASHI, 2015).

A classificação distribui os periódicos em estratos: A1 (o mais elevado), A2, B1, B2, B3, B4, B5, e C (este com peso zero). No entanto, um mesmo periódico pode figurar em estratos diferentes, dependendo da área de conhecimento. (VOGUEL E KOBASHI, 2015).

Para Furnival (2015) a questão do desempenho da Via Verde no Brasil, envolve o processo de avaliação da produção dos programas de pós-graduação. Para a autora, os pesquisadores brasileiros têm preferência por publicar em periódicos com alto fator de impacto, pois além do fato de serem considerados mais prestigiados, é condição para obter financiamentos e verbas de órgãos públicos de fomento, que dá enfoque na avaliação utilizando dados quantitativos.

Os participantes comentaram sobre o peso do fator de impacto na avaliação da produção dos programas de pós-graduação pela Capes. O órgão leva em conta o fator de impacto dos periódicos para classificá-los no sistema Qualis. Com isso, pesquisadores e suas instituições com publicações nessas revistas são melhores avaliados e, portanto, podem ser beneficiados por financiamentos públicos. (FURVINAL, 2015).

Outra característica que apontamos nesta pesquisa como sendo um fator que pode afetar o desenvolvimento do AA no Brasil é a Comunicação da Ciência nacional. Para alguns autores como Spagnolo (1989); Guédon (2010), o Brasil, assim como alguns países da América Latina, são considerados países fazedores de Ciência Periférica, o que lhe atribui características específicas que afetam a Comunicação da ciência no país.



Para Spagnolo (1989) a investigação científica no Brasil está fortemente concentrada em Universidades e intimamente associada aos programas de pós-graduação. Em entrevista realizada com pesquisadores brasileiros, Spagnolo (1989) apontou três questões, indicadas pelos pesquisadores, como motivos para pesquisadores do Terceiro Mundo publicar em periódicos internacionais, a saber: a) maior visibilidade oferecida por periódicos internacionais e, conseqüentemente, a oportunidade de mais prestígio; b) pressão indireta e até mesmo direta de agências governamentais, onde se você não mostrar a produção internacional, é muito mais difícil de obter apoio para o seu pesquisar projetos e c) Não existe um fórum nacional adequado para o debate em algumas áreas, como por exemplo a Química.

Isto posto, podemos inferir que existe uma exigência por parte de agências de fomento para se publicar internacionalmente. Este quesito fica ainda mais clara em áreas como a Física, Engenharias, Ciências da Saúde, etc, que se constituem áreas mais internacionalizadas. Ou seja, há uma influência governamental para promoção da comunicação científica nacional.

Como o AA é apontado como o caminho mais democrático para se obter acesso às publicações científicas, possivelmente esferas governamentais apoiam o desenvolvimento do AA através de infraestrutura tecnológica (desenvolvimento de sistemas de gerenciamento de periódicos on-line e repositórios), financiamento de projetos que visam o AA, etc.

Guédon (2010) indica que o modo de comunicar da Ciência um fator que influencia o desenvolvimento de políticas em AA num país. Para o autor há uma distinção entre os países de Ciência Central (*mainstream*)<sup>6</sup> dos países Ciência Periférica, no que tange a Comunicação Científica e existindo também à possibilidade desta convergência no AA.

Quando falamos em Ciência Central ou Mainstream Science nos referimos a divisão do mundo pelo paradigma científico atual que teve origem na Europa no século 15 e 16, evento que ocorreu juntamente com a expansão colonial, todavia a Ciência Central obteve maior amplitude após as Revoluções Industrial e Científica.

A ciência central, ou “mainstream science”, constitui o paradigma científico vigente, inicialmente na Europa, estendendo-se aos poucos a todos os continentes, até sua completa mundialização: esta ciência apresenta características nítidas, como uma hierarquização e uma ligação estreita com os interesses econômicos dominantes. Sua prática pressupõe atualmente várias regras, entre as quais podem apontar-se: trabalho em equipe, cooperação e colaboração, divulgação ampla dos resultados em uma língua franca internacional (que tem variado com o tempo, do latim ao francês ou alemão, agora o inglês), avaliação constante pelos demais praticantes da ciência, e alguma ligação com as necessidades econômicas e sociais da sociedade onde ela se desenvolve. (FILGUEIRAS, 2001, grifo nosso).

<sup>6</sup>Constitui um paradigma científico vigente pela hierarquização dos países ditos de primeiro mundo com interesses econômicos dominantes.



O periódico científico surgiu no século 17 com a premissa de transmitir de forma rápida os estudos e descobertas realizadas. (Meadows, 1999). Quando o periódico nasceu, já havia a divisão do paradigma científico em Ciência Central e Periférica. Assim o meio mais antigo e tradicional de comunicação da Ciência já nasceu num ambiente pré-estabelecido de dominação europeia. Fator este que pode ter acentuado a desigualdade na divisão da ciência entre os países economicamente hegemônicos e aqueles países colonizados ou economicamente debilitados.

Os periódicos científicos surgiram na segunda metade do século XVII devido a várias razões. Algumas eram específicas (como a expectativa de seus editores de que teriam lucro); algumas gerais (como a crença de que para fazer novos descobrimentos era preciso que houvesse um debate coletivo. O motivo principal, contudo, encontra-se nessa necessidade de comunicação, do modo mais eficiente possível, com uma clientela crescente interessada em novas realizações. (MEADOWS, 1999, p. 7).

Assim, do mesmo modo que o êxito dos países em progredir os Periódicos Científicos está estritamente ligado ao episódio do desenvolvimento econômico, podemos inferir que o modo de desenvolvimento do acesso aberto nos países está diretamente ligado à divisão da ciência no mundo em função do paradigma científico. Para Guédon, o Movimento do Acesso Aberto tem importância, pois:

[...] promete fazer o sistema mundial da ciência funcionar melhor, ao aprimorar a infraestrutura de comunicação. E, mais importante, o AA abre expectativas plenas de esperança. Isto é, como o Acesso Aberto, corrigir muitas das injustiças inerentes à barreira de conhecimentos existentes entre os chamados países centrais e periféricos torna-se uma possibilidade inegável. (GUÉDON, 2010, p. 73).

Assim como Guédon (2010), Ortellado aponta a influência da divisão da ciência (Central e Periférica) na adoção de medidas para o AA e reconhece a Via Verde como garantia de acesso aos artigos de periódicos brasileiros produzidos com recursos públicos.

Segundo Terra-Figari (2008), os países da “periferia científica” são prejudicados por desvantagens em tamanho da comunidade acadêmica (que é muito pequena para se produzir uma massa crítica adequada); incentivo (as publicações não contam muito na hora das promoções); financiamento (a ciência é uma atividade marginal e não tem papel de destaque como nos centros da ciência, é realizada em instituições públicas e tem como fonte de financiamento, Organizações Não Governamentais (ONGs) e fundações internacionais), linguagem de publicação (inglês é o idioma universal) e edição de parâmetros (normatização e tipo de estudos, como por exemplo a falta de artigos originais).

Para Vessuri (2015) o que determina a baixa participação da América Latina na produção mundial não é somente a ciência não ser mainstream. A exclusão das revistas científicas latino americanas das principais bases de dados de indexação e a barreira da língua interferem em seu progresso científico. Esses fatores, assim como o baixo investimento em inovação e desenvolvimento, impedem que haja uma competição igua-

tária entre os países ditos periféricos.

Segundo Alperin, Fischman e Willinsky (2008) as motivações da América Latina para promoção do Acesso Aberto à produção científica são diferentes das motivações dos países centrais. Enquanto na Europa, Estados Unidos e Canadá, o AA tem sido impulsionado em parte, por um aumento nos preços e na quantidade de periódicos (cri-se dos periódicos), as publicações na América Latina têm sido tradicionalmente livres.

Mas para Guédon (2010) o Brasil ser considerado um país de ciência periférica foi fator determinante para o maior desenvolvimento na Via Dourada. Ortellado (2008) também considera a característica de Nação Periférica o fato de o Brasil disponibilizar a maior parte de seus periódicos de excelência na Web.

As condições nas quais essa experiência se desenvolveu, no entanto são particulares de uma nação periférica. Na ausência de um relevante mercado nacional de revistas científicas, os periódicos brasileiros foram historicamente financiados quase que exclusivamente com recursos públicos. Neste cenário, a conversão para um modelo de acesso aberto foi apenas uma questão de deliberação políticas das agências financiadoras. (ORTELLADO, 2008).

Para Costa e Leite (2016) o desenvolvimento e consolidação do AA “vem ocorrendo de forma diferente nos países desenvolvidos e do resto do mundo, particularmente na América Latina”. Para os autores, esta diferença seria pelas características do sistema de comunicação científica estabelecida na América Latina ao longo do tempo. E apesar da consolidação do AA Dourado na América Latina, devido à natureza não-comercial da maioria dos editores científicos na região, ser referência, não foi suficiente para impulsionar o AA Verde.

## **Materiais e Métodos**

Esta pesquisa é cunho descritivo e de abordagem quantitativa que se dará por meio da análise da produção científica das IES, no caso os artigos de periódicos, do ano de 2013, dos cursos de pós-graduação nas áreas de: Ciências Exatas (CE) representada pela Física, Ciências Sociais (CS) representada pela Economia e Ciências Humanas (CH) representada pela Psicologia, inseridas na Plataforma Sucupira.

A seleção dos casos foi baseada no *ranking Web of Repository* em que os repositórios brasileiros estavam nas primeiras colocações em 2013. Os Programas foram selecionados tendo em vista as áreas temáticas dos repositórios pioneiros: Física (*arXiv*), Psicologia (*CogPrints*) e Economia (*RePEC*). A escolha das áreas está baseada no artigo de Ortellado (2008) que faz uma distinção na aceitação do AA na área de Ciências Naturais e da Terra nas demais áreas tais como as Ciências Sociais e Humanas. A produção científica analisada somou 525 artigos.

Os respectivos RIs serão consultados para verificar se os artigos levantados na plataforma sucupira foram depositados nos RIs das IES.

Os artigos serão elencados por tipo de acesso nas categorias elaboradas segundo Harnad (2012) que são: Acesso Restrito (Não AA); Acesso Aberto através de RI (Acesso Aberto Verde); Acesso Aberto no site do periódico e/ou em base de dados (Acesso Aberto Dourado).

Quadro 1: Categorias por Tipo de Acesso

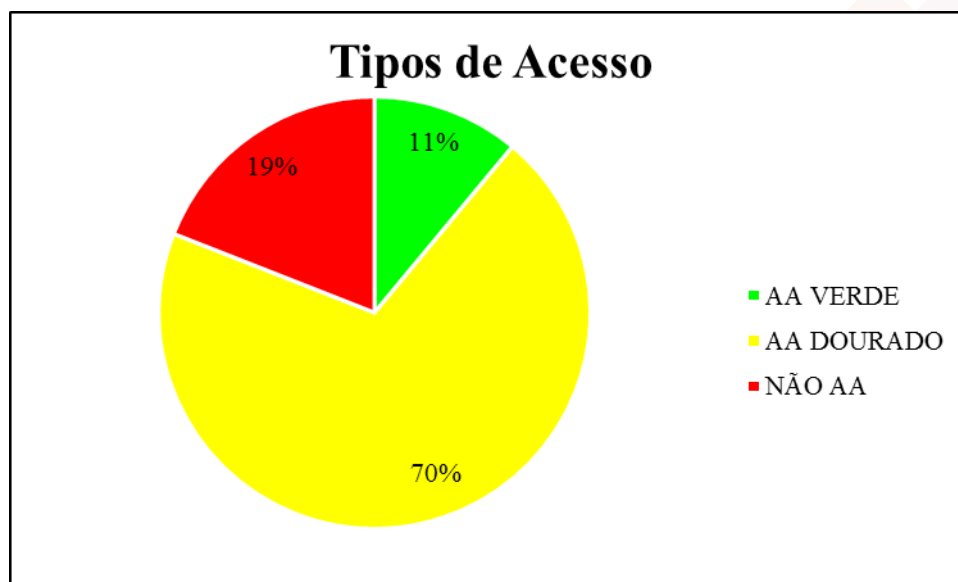
CATEGORIA	TIPO DE ACESSO	DESCRIÇÃO
Acesso Restrito	Não AA	O artigo é restrito ao usuário sendo disponível somente através da assinatura do periódico e não está disponível em RI
Acesso Aberto através de RI	AA Verde	O artigo tem acesso restrito ao usuário no periódico comercial, mas é disponibilizado livremente em RI
Acesso Aberto no site do periódico e/ou em base de dados	AA Dourado	O artigo é disponível ao usuário no site do periódico e/ou base de dados como SciELO, Pepsic, etc.

Fonte: O autor (2015).

## Resultado Final

A pesquisa mostrou que a maioria dos artigos de periódicos presentes nos RIs analisados e no período delimitado, são provenientes da via dourada, totalizando 70%. Por outro lado, 19% da produção dessas áreas não estão disponíveis nos RIs e seu acesso está bloqueado exigindo pagamento (Não AA). Apenas 11% dos artigos são provenientes da via verde sinalizando, em linhas gerais, uma baixa adesão a essa estratégia em relação à via dourada.

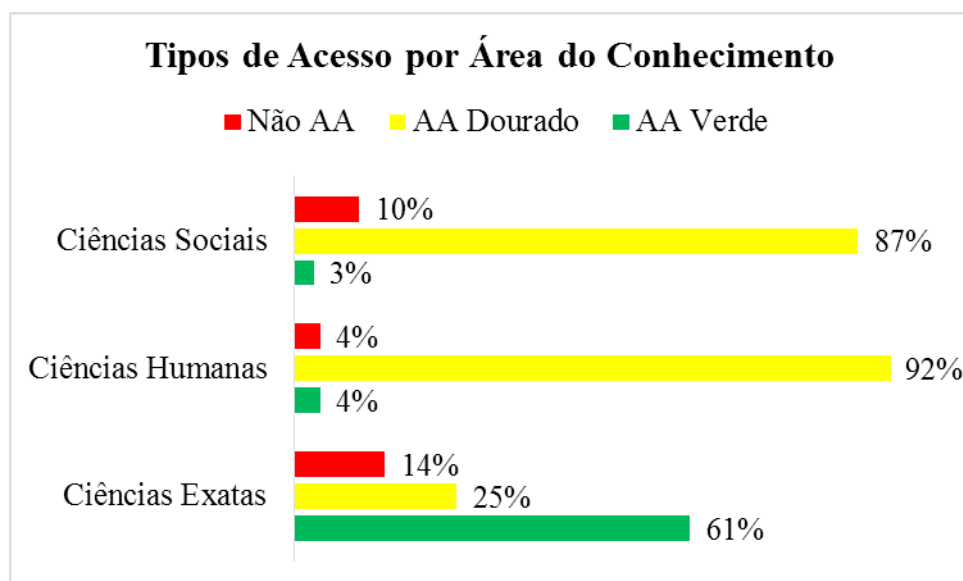
Gráfico 2: Tipos de Acesso em RIs no Brasil



*Fonte: O Autor (2015).*

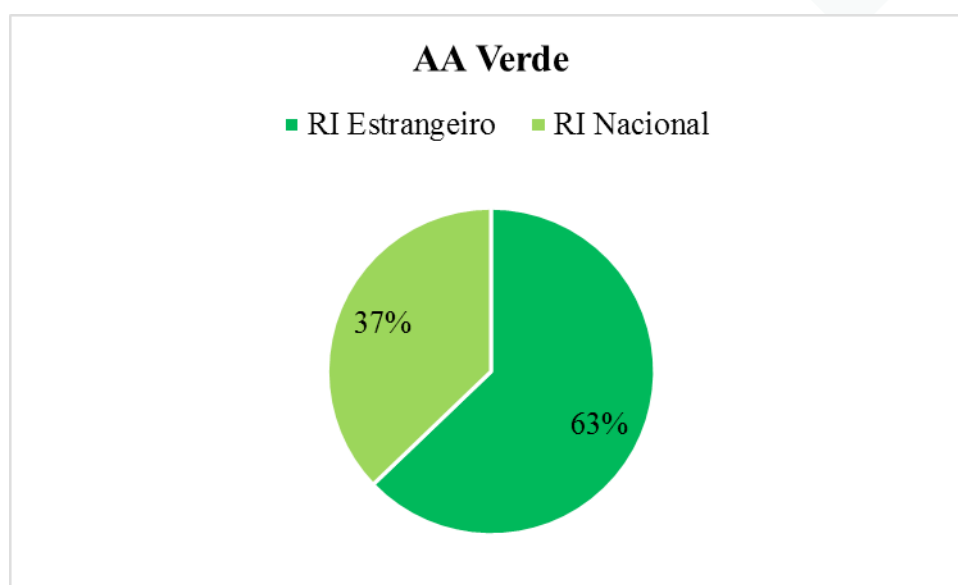
Quando o foco da análise envolve as áreas estudadas, são notórias algumas particularidades: a CE possui porcentagem equilibrada de artigos sob AA Verde (34%) e sob o AA Dourado (27%), no entanto, demonstrou ser a área que mais necessita do AA Verde por possuir a maior proporção de Não AA (38%) na produção científica; a CH destacou-se no AA Dourado (89%), com pouco desempenho no AA Verde; já a CS demonstrou um desempenho nulo no AA Verde, predominância do AA Dourado (92%) e 16% dos artigos científicos Não AA.

Gráfico 3: Tipos de acesso por Áreas do conhecimento



*Fonte: O autor (2014).*

Gráfico 4: Acesso Aberto Verde por Repositórios



*Fonte: O Autor (2016).*

Com base na produção científica acessível através do AA Verde foi possível ainda verificar através desta amostragem que 63% dos artigos de publicados por autores bra-

sileiros em IES nacionais, estão disponíveis em AA Verde por meio de RI estrangeiros, como por exemplo no *arXiv*, em RI de Universidades estrangeiras, através de parcerias internacionais dos nossos autores e em sites como Academia e ResearchGate.

## Conclusões

Mapeando a produção científica das IES publicada em periódico e o arquivamento destas publicações em RIs, podemos notar uma baixa adesão do AA Verde, cerca de 11% e uma valorização do AA Dourado, cerca de 70%. É notável também que parte da produção científica é restrita, aproximadamente 19% Não AA, tendo destaque para o Acesso Aberto Dourado, 60%, o que nos leva a constatar que o RIs nacionais tem pre-domínio de depósito de Artigos publicados em Periódicos de AA.

A categorização dos artigos de periódicos por tipo de acesso indica que nas CH e CS há maior sensibilidade às iniciativas nacionais de promoção do AA investindo em periódicos AA, apresentando relevância na AA dourado. Nas CE, apesar de semelhante desempenho no AA verde e AA dourado, os dados apontam que há maior necessidade da AA verde, 38% Não AA, uma vez que publicam mais em periódicos comerciais, possuindo porcentagem considerável de produção científica sob a AA verde, 34%. A CH demonstrou maior sensibilidade no AA Dourado, cerca de 90% e uma baixa adesão ao AA Verde, 1%. A CS também demonstrou considerável desenvolvimento em AA Dourado, 84% e AA Verde inexistente.

Verificando a produção de artigos científicos de IES brasileiras inseridas em RIs nacionais, há apontamentos que sugerem características particulares, com ênfase na AA dourado, como o predomínio de depósito de artigos de periódicos AA, característica essa que não é inerente à AA verde apontada pela literatura e pelos signatários em AA e praticada em grandes instituições de ensino e pesquisa no mundo, onde se prioriza o depósito em RIs, de pesquisa aprovada em periódicos comerciais.

Esta pesquisa torna premente a necessidade de verificar em novas investigações se essas características estão presentes em outras áreas do conhecimento e em demais instituições de ensino e pesquisa e em seus respectivos RIs, visando um panorama completo do AA Verde no Brasil e se há outras características não contempladas neste estudo.

## Referências

Budapest Open Access Initiative. (2002). Iniciativa de Budapeste pelo Acesso Aberto. Budapeste. Disponível em: <<http://www.budapestopenaccessinitiative.org/translations/portuguese-translation>>. Acesso em: Março 2014.

Furnival, A. C., Oprime, P. & Silva Jerez, N. S. (2013). Um estudo da percepção do acesso aberto à literatura científica por pesquisadores brasileiros. 4ª CONFERÊNCIA LUSO-BRASILEIRA SOBRE ACESSO ABERTO – CONFOA, São Paulo.

Steven, H. Brody, T. Vallières, F. Carr, L. Hitchcock, S. Gingras, Y.,... (2004). The green and the gold roads to Open Access. Nature Web Focus. Disponível em: <<http://eprints.soton.ac.uk/259940/1/21.html>>. Acesso em: Março 2014.

Steven, H. (2012). Why the UK should not heed the Finch Report. Impact of Social Sciences blog. Disponível em: <<http://blogs.lse.ac.uk/impactofsocialsciences/2012/07/04/why-the-uk-should-not-heed-the-finch-report/>>. Acesso em: Dez. 2014.

Ortellado, P. (2008). As políticas nacionais de acesso à informação científica. Liinc em revista. Rio de Janeiro, 4 (2), p. 186-195.

Weitzel, S. (2014). As novas configurações do acesso aberto: desafios e propostas. Revista Eletrônica de Comunicação, Informação & Inovação em Saúde, Rio de Janeiro, 8 (2) p.65-75. Disponível em: <<http://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/447/1102>>. Acesso em: Jan. 2015.

# “Módulo de Revisión por Pares en Abierto para Repositorios de Acceso Abierto”

*Pandelis Perakakis (1), Agnès Ponsati (2), Isabel Bernal (3), Carles Sierra (4), Nardine Osman (5), Concha Mosquera-de-Arancibia (6), Emilio Lorenzo (7)*

---

## RESUMEN

Presentamos el primer módulo de revisión por pares abierto para repositorios de acceso abierto. El módulo, diseñado en esta primera etapa para la integración con repositorios DSpace, para evitar ‘endogamia’, permite a cualquier especialista ofrecer una evaluación cualitativa y cuantitativa de cualquier trabajo de investigación alojado en un repositorio compatible. El sistema se apoya en el uso de métricas ponderadas de reputación para artículos, revisiones, autores y revisores. Una función de búsqueda avanzada permite a los usuarios del repositorio filtrar u ordenar los trabajos de investigación por su reputación, que se calcula basándose en las revisiones que recibe. La integración de la revisión por pares en los repositorios promueve la discusión abierta al permitir una colaboración directa, abierta y transparente entre los autores y los revisores, y produce nuevas métricas que reflejan directamente la calidad de un trabajo de investigación percibida por los colegas expertos, al contrario que las métricas actuales disponibles que dan cuenta de la calidad solo de manera indirecta a través de las estadísticas de uso. El sistema de revisión por pares implementado permite que los revisores de un trabajo y los autores del mismo (derecho de réplica) que estén en la base de datos del repositorio pueden hacer comentarios sobre las revisiones del trabajo. Esta es una opción que podría ser ampliada a todos los especialistas en la materia registrados en el repositorio pero sin tenerlos en cuenta en las métricas. El módulo de revisión por pares en abierto ya se ha instalado en dos importantes repositorios españoles (DIGITAL.CSIC, e-IEO) con resultados iniciales prometedores.

**Palabras clave:** Revisión por Pares en Abierto; Métricas; Reputación; DSpace; DIGITAL.CSIC; e-IEO.

## ABSTRACT

We present the first open peer review module for open access repositories. The module, designed in this first stage for integration with DSpace repositories, in order to avoid “inbreeding”, enables any scholar to offer a qualitative and quantitative evaluation of any research object hosted in a compliant repository. Weighted reputation metrics are calculated for articles, reviews, authors and reviewers. An advanced search function allows repository users to filter or sort research objects based on their reputation, which is calculated based on the reviews received. The integration of peer review in repositories



promotes open scholarship by enabling a direct, open and transparent collaboration between authors and reviewers, and produces novel metrics directly reflecting the perceived quality of a research work by expert peers, contrary to current available metrics that only indirectly account for quality through usage statistics. Reviewers of the work and the authors of the work (right of reply) at the repository data base can comment on the reviews of that work. This option may be extended to all specialists in the field registered at the repository but without taking them into account in metrics. The open peer review module has already been installed in two major Spanish repositories (DIGITAL.CSIC, e-IEO) with promising initial results.

**Keywords: Open Peer Review; Metrics; Reputation; DSpace; DIGITAL.CSIC; e-IEO.**

## **INTRODUCCIÓN**

La investigación tradicional se está transformando en una dinámica interconectada en la que ideas, herramientas y resultados pueden ser accesibles al instante a toda la comunidad científica. Los repositorios de acceso abierto juegan un papel crucial, permitiendo un acceso inmediato a todo tipo de resultados de investigación, pero un elemento importante que aún falta en ellos es una evaluación tanto cualitativa como cuantitativa de los artículos de investigación que facilite la selección del contenido más relevante. Las métricas disponibles actualmente (número de visitas y descargas) no reflejan la calidad de un trabajo, que solo puede ser determinada directamente por revisores que ofrezcan su opinión como expertos, junto con puntuaciones cuantitativas basadas en criterios específicos.

Para solucionar este problema, desarrollamos un Módulo de Evaluación por Pares en Abierto (OPRM: *Open Peer Review Module*) para repositorios, ofreciendo este servicio añadido. De esta forma, cualquier trabajo de investigación (preprint o ya publicado) puede ser evaluado por un número ilimitado de expertos que ofrecen una revisión cualitativa en forma de texto y medidas cuantitativas que construyen la reputación del trabajo de investigación y de sus autores.

Este sistema de evaluación es abierto y transparente. En nuestro modelo, abierto (los textos completos de las revisiones están públicamente disponibles junto con el trabajo evaluado) y transparente (se revela la identidad de los revisores) son dos aspectos necesarios para solucionar sesgos inherentes al sistema anónimo de evaluación por pares en el que los evaluadores no se responsabilizan públicamente de sus revisiones.

La colaboración y la discusión abierta y transparente entre pares expertos, donde todas las interacciones entre autores y revisores van firmadas, tienen como objetivo el incremento de la calidad científica de la obra.

## **MATERIALES Y METODOLOGÍA**

### **Cómo funciona el OPRM**

- En el OPRM, ya implementado en dos importantes repositorios españoles (DIGITAL.CSIC, e-IEO), técnicamente, cualquier especialista puede solicitar la evaluación de un trabajo, aunque, de momento, en los dos repositorios citados, solo pueden solici-

tar revisiones de un trabajo:

- » En el flujo de archivo del trabajo en el repositorio: el autor o el administrador de colección que hace el envío.
  - » Una vez archivado: el administrador general del repositorio y los autores del trabajo que estén en la base de datos del repositorio.
- Para evitar ‘endogamia’, cualquier especialista en la materia puede evaluar un trabajo, esté o no en la base de datos del repositorio.
  - Los revisores de un trabajo pueden comentar las otras revisiones del trabajo y los autores del mismo que estén en la base de datos del repositorio pueden comentar las revisiones de su trabajo.
  - La solicitud se hace pulsando “Revisar este trabajo” en la página del ítem.
  - Los revisores reciben la solicitud de evaluación por correo electrónico con un enlace a un sencillo formulario (Figura 1) donde:
    - Rellenan: nombre, afiliación, idioma.
    - Proporcionan cuatro calificaciones: impacto del trabajo en su campo, en otras áreas, para la sociedad y valoración global.
    - Deciden si alcanza el estándar científico.
    - Dan su opinión sobre el trabajo. Esta opinión se puede adjuntar también en pdf.
    - La revisión se publicará con licencia CC-By.
  - Cualquier comentario sobre una revisión se puede hacer utilizando un formulario similar, con una única valoración global, pulsando “Comentar esta revisión”.
  - El administrador del sistema recibe en su workflow las revisiones que, tras su aprobación, se enlazan al trabajo evaluado y se hacen públicamente accesibles.

### **Principales aspectos de la instalación del OPRM**

- Tanto e-IEO (DSpace v. 5.2, XMLUI) como DIGITAL.CSIC (DSpace-CRIS v. 4.3, JS-PUI) (Figura 2) cuentan con modelos de autor avanzados necesarios para:
  - » Desambiguar e identificar inequívocamente a los autores.
  - » Mostrar en las páginas de autor las reputaciones calculadas por el OPRM.
- Los submódulos de invitaciones y revisiones se han desarrollado como una extensión del flujo de envío de trabajos de DSpace, mediante configuraciones estándar a los formularios de envío y flujos de revisión de DSpace.
- El código se puede descargar del repositorio de software del proyecto. Su instalación y aplicación se realizan como cualquier otro módulo o add-on de DSpace.
- El modelo de datos de DSpace se extiende para dar cabida a las puntuaciones, valoraciones y otras entidades del sistema OPRM.
- La visualización de ítems se modifica para mostrar las reputaciones y relaciones de objetos, revisiones y comentarios. La visualización de las páginas de autor se modifica para incluir las reputaciones de los autores, aspecto clave de la difusión.
- Recomendamos también ajustes adicionales al sistema de búsqueda, índices y filtrado, con el fin de que los usuarios puedan identificar los objetos con revisiones, los de mayor reputación, etc.
- Se deben efectuar ajustes a la configuración de OAI-PMH, mediante parámetros de filtrado, con el fin de no exponer por la interface la información de revisiones.

Disponibilidad del código

- Código para DSpace v. 5 XMLUI (e-IEO):  
<https://github.com/arvoConsultores/Open-Peer-Review-Module>
- para DSpace-CRIS v. 4 JSPUI (DIGITAL.CSIC):  
<https://github.com/arvoConsultores/dspace-cris>
- Wiki del proyecto:  
<https://github.com/arvoConsultores/Open-Peer-Review-Module/wiki>

## El modelo de evaluación de la reputación

Se basa en la evaluación cuantitativa de los trabajos y de los comentarios de estas evaluaciones hechos por los pares. Es importante destacar que nuestro módulo incluye un sistema de reputación del revisor basado en la evaluación de sus revisiones por otros revisores. Esto permite un escalado sofisticado de la importancia de cada revisión en la evaluación final de un trabajo, basado en la reputación del revisor.

El modelo de revisión permite la calificación del impacto del trabajo en su campo, en otras áreas y para la sociedad pero es necesaria una valoración global para calcular la reputación del trabajo (Osman y otros, 2014); (Sabater-Mir & Sierra, 2005).

El modelo (Osman & Sierra, 2016) cuantifica:

- Reputación del trabajo: una agregación ponderada de las calificaciones globales de sus revisiones, donde el peso es la reputación como revisor del investigador que la realizó. La reputación del trabajo no depende de una única revisión; se necesitan dos. Cuando no hay suficientes revisiones, la reputación del trabajo está sin definir. (Figura 3).
- Reputación como autor: una agregación de las reputaciones de sus trabajos. El impacto de la reputación de un trabajo particular en la reputación general como autor es inversamente proporcional al número de autores del trabajo. (Figura 4).
- Reputación de la revisión: una agregación ponderada de las calificaciones recibidas en los comentarios, donde el peso es la reputación como revisor del investigador que escribió el comentario. La reputación de una revisión no depende de un único comentario, son necesarios dos. Cuando no hay suficientes comentarios, la reputación de una revisión se considera igual a la reputación como autor del revisor. (Figura 5).
- Reputación como revisor: la reputación de un investigador como buen revisor. Se obtiene de los comentarios sobre sus revisiones y de la similitud entre comentarios (si mis comentarios están próximos a los comentarios de otro investigador, entonces nuestras reputaciones como revisores están próximas). Antes de que se recojan suficientes comentarios, la reputación como revisor de un investigador se asume que es la misma que su reputación como autor. Se adjudica 50/100 (es decir, no se conoce) si todavía no existe tal reputación. (Figura 6).

Las revisiones y los comentarios, con sus métricas, se recogen en las colecciones de la comunidad OPRM. (Figura 7).

Conviene hacer notar que cuanto mayor sea el número de trabajos revisados de cada autor y mayor el número de revisiones y comentarios de cada trabajo, más “exactas” serán las métricas de las reputaciones.

Es importante tener en cuenta que los algoritmos de reputación se desarrollaron como plugins por separado, lo que permite adaptarse fácilmente a otros modelos de reputación. Por tanto, es factible su adaptación funcional a otros repositorios.

## El proyecto piloto

Una vez instalado el módulo, se planteó un proyecto piloto para probar el módulo en ambos repositorios.

**En e-IEO** se contó con la colaboración de nueve científicos del IEO:

- Tres autores para ser evaluados.
- Uno por área de investigación del IEO (Pesquerías, Acuicultura, y Medio Marino y Protección Ambiental).
- Tres trabajos, ya publicados, para ser evaluados (uno por autor).
- Seis revisores para hacer revisiones y comentarios.
- Los autores del trabajo del IEO también pudieron comentar las revisiones de sus trabajos.

**En DIGITAL.CSIC** se lanzó una campaña de captación de un grupo selecto de investigadores de diversas áreas científicas: Ciencias Químicas, Humanidades y Ciencias Sociales, Biomedicina, Biología, Recursos Naturales, Ciencias Físicas, y Ciencias y Tecnologías de Ciencia de la Alimentación. El equipo del repositorio recurrió a la plataforma de investigadores Publons para identificar autores CSIC sensibilizados con el tema de las revisiones y evaluación científicas y a favor de una reforma del sistema. Además, se contactó con investigadores CSIC muy activos y concienciados con el acceso abierto.

- DIGITAL.CSIC fomentó la revisión externa, no institucional, pero fueron los autores los que decidieron libremente qué trabajos elegir para pedir revisiones abiertas y los nombres de los expertos a invitar.
- Tres investigadores CSIC no han recibido aún las revisiones solicitadas a través del módulo.
- Fueron evaluados tres trabajos (dos artículos, ya publicados, sobre Recursos Naturales y Ciencias Sociales, respectivamente, y una comunicación a congreso, sobre investigación en Robótica e Ingeniería).
- Cada artículo científico recibió dos revisiones: mientras que en el caso de *Towards a unified paradigm for sequence-based identification of fungi* los revisores fueron externos al CSIC (concretamente, de los National Institutes of Health de los Estados Unidos), el artículo *Bibliometric and benchmark analysis of gold Open Access in Spain: big output and little impact* fue evaluado por dos investigadores del CSIC. La comunicación de congreso *A New Beamforming Process Based on the Phase Dispersion Analysis* ha sido evaluada por un solo experto, de la institución brasileña UNESP.
- Los autores de los tres trabajos participantes en este piloto valoraron como muy útiles las revisiones recibidas: los tres coincidieron en la calidad y extensión de las revisiones y les resultaron útiles. Por el momento, solo la autora del artículo *Towards a unified paradigm for sequence-based identification of fungi* ha comentado las dos revisiones recibidas.

## RESULTADOS

En e-IEO se obtuvieron:

- Seis revisiones, dos por trabajo.
- Doce comentarios, dos comentarios por revisión: uno del otro revisor del trabajo y el otro del autor del IEO.
- Y métricas de reputación ponderadas para trabajos y revisiones (enlazadas con cada trabajo) y para autores y revisores (reflejadas en los perfiles de autor).

Hay que tener en cuenta que, en este estudio, la reputación como autor se ha calculado a partir de las revisiones sobre un único trabajo.

En el repositorio, la página de un trabajo muestra el valor de la reputación del trabajo (Reputación de la publicación), los enlaces a sus revisiones y los valores de calificación dados por cada revisor (Figura 8).

La página de una revisión presenta la revisión (en pdf), la valoración global asignada por este revisor al trabajo, la reputación de esta revisión, el enlace al trabajo evaluado (con su valor de reputación) y los enlaces a los comentarios hechos a esta revisión (por otro revisor y por el autor), con sus correspondientes calificaciones.

La página de un comentario muestra el comentario, la valoración global asignada por el autor de este comentario (un revisor o el autor) a la revisión que comenta, el enlace a esa revisión con su valor de reputación. Ver enlaces de trabajos evaluados.

Las reputaciones como autor y como revisor se muestran en las páginas de perfil de autor (Figura 9).

- *Trabajos evaluados con sus métricas de reputación*
- <http://hdl.handle.net/10508/8123> (Pesquerías: artículo publicado en *Journal of Marine Biology*, 2011).
- <http://hdl.handle.net/10508/2494> (Acuicultura: póster, resumen publicado en *Aquaculture Europe 14 Congress*, 2014).
- <http://hdl.handle.net/10508/7818> (Medio Marino y Protección Ambiental: artículo publicado en *Marine Ecology Progress Series*, 2009).

*Ejemplos de perfiles de autor con sus reputaciones como autores y revisores (ver “Mostrar detalles”)*

- de-la-Gándara, F. (Fernando).
- García-Rodríguez, M. (Mariano).
- Jerez, S. (Salvador).
- Orejas, C. (Covadonga).
- Rodrigues-dos-Santos-Domingues, P.M. (Pedro Miguel).

## *Opiniones de los autores y revisores invitados al proyecto piloto*

- El OPRM era un servicio esperado que conduce a una colaboración entre pares.
- Al conocer la identidad de los revisores, los autores pueden estar seguros de que las evaluaciones han sido hechas por expertos.
- El intercambio de información (los textos de revisiones y comentarios) entre expertos evita una posible subjetividad de los revisores y facilita la autoevaluación de los revisores.
- Pero, el actual sistema de evaluación por pares de las revistas está tan integrado en la comunidad científica que la implementación del OPRM puede requerir tiempo.
- Algunos investigadores sugieren que una evaluación negativa podría crear una situación incómoda entre colegas.
- Sin embargo, muchos investigadores esperan que el pleno apoyo al OPRM conduzca a la ciencia completamente abierta a través de plataformas de evaluación con métricas independientes comparables.

**En DIGITAL.CSIC** se obtuvieron:

- Cinco revisiones de tres trabajos (dos artículos publicados y una comunicación a congreso).
- Dos comentarios.
- Y métricas de reputación ponderadas para trabajos (enlazadas con cada trabajo) y para autores y revisores (reflejadas en los perfiles de autor). La configuración del módulo de reputaciones de autores y revisores está en fase de ajuste y se explica más adelante.

Los trabajos evaluados presentan, en metadatos específicos del módulo, enlaces a las revisiones junto con sus respectivas valoraciones globales (de cada revisor).

El módulo OPRM permite realizar revisiones extensas ya que posibilita adjuntar tantos ficheros como sea necesario para respaldarlas (Figura 10). Los comentarios realizados por los autores de los trabajos revisados usan esta funcionalidad para agradecer, comentar y valorar las evaluaciones recibidas (Figura 11).

Trabajos evaluados con sus métricas de reputación

- <http://digital.csic.es/handle/10261/131134>
- <http://digital.csic.es/handle/10261/130958>
- <http://digital.csic.es/handle/10261/44699>

Las reputaciones como autores y revisores aparecerán en los perfiles personales una vez que se hayan pasado a producción unos ajustes mediante los cuales la reputación se calcula tomando como base:

- Al menos dos revisiones recibidas.
- El trabajo revisado exclusivamente, y no ponderado, como se ha hecho en una primera instalación, con el volumen de trabajos del mismo autor que no han recibido ninguna revisión a través del módulo. Este ajuste ha sido necesario porque en la primera instalación se penalizaba a los autores con muchos trabajos en el repositorio.



## *Opiniones de los autores y revisores invitados al proyecto piloto*

- Un servicio innovador y largamente esperado en el repositorio.
- Su potencialidad será mucho mayor cuando no sea necesario invitar formalmente a ningún experto.
- En algunas disciplinas muy restringidas, en que toda la comunidad científica se conoce, implantar este sistema de revisiones presenta muchos retos.
- Qué funcionalidades adicionales presenta el módulo para que pueda competir con los sistemas de evaluación y reputación abiertas relativamente consolidados como los de ResearchGate y Academia.edu.
- La utilidad para libros y monografías no es evidente pero su uso sistemático en artículos preprints es un importante reto y tiene también potencial para otros resultados de investigación (datasets, programas informáticos, material divulgativo y docente, informes, propuestas de contribuciones a congresos, conferencias,...).
- Hacer buenas revisiones requiere tiempo y esta dedicación extra necesita una motivación sólida ligada al reconocimiento institucional y en CV. Por esto, es necesario implantar un sistema de recompensa/reconocimiento para los investigadores. Además, los investigadores deben sentir que tienen algo que aportar realmente.

## **CONCLUSIONES**

Los repositorios institucionales de acceso abierto desempeñarán un papel mucho más importante en la comunicación entre investigadores mediante la integración de un sistema de evaluación abierto y transparente que se ofrece como un servicio de valor añadido. Esto puede ayudar a resolver muchos de los problemas relacionados con el sistema actual de reputación de la investigación basado en las revistas, así como ofrecer a los investigadores más incentivos para colocar sus resultados en repositorios donde puedan ser evaluados en abierto por pares expertos. Hemos desarrollado el primer módulo de revisión por pares en abierto para los repositorios utilizando DSpace, pero el módulo se puede adaptar fácilmente para su uso con otros paquetes de software. El módulo presenta un nuevo modelo de reputación que mejora significativamente la calidad de la evaluación de los trabajos de investigación y de sus autores. Su uso sistemático en preprints es un importante reto y conducirá a la ciencia completamente abierta.

## **BIBLIOGRAFÍA**

Osman, N., Provetti, A., Riggi, V., & Sierra, C. (2014). MORE: Merged Opinions Reputation Model. En: Proceedings of the 12th European Workshop on Multi-agent Systems (EUMAS 2014). Springer.

Osman, N., & Sierra, C. (2016). Reputation in the Academic World. En: J. Zhang, R. Cohen y M. Sensoy (eds.). Proceedings of the 18th International Workshop on Trust in Agent Societies, 09-mayo-2016. Singapore: <http://ceur-ws.org>.  
<http://digital.csic.es/handle/10261/130842>.

Sabater-Mir, J., & Sierra, C. (2005). Review on computational trust and reputation models. Artificial Intelligence Review, 24, 33-60.

Escriba su opinión sobre el trabajo incluyendo los detalles necesarios para ayudar a los autores a mejorar su contribución. Procure ser constructivo. Su revisión será publicada con el trabajo original bajo una licencia CC-BY: uso público, siempre que se reconozca su autoría mediante una cita del formato: Revisor. Fecha. Título del trabajo original [Revisión]. Autores del trabajo. [Handle/ur](#).

**Autor: \***  
 Apellidos: si se utilizan dos apellidos, irán unidos por guión, p. ej. *Pérez-de-Gómez; Fernández-Martín*. Introduzca uno de los apellidos para buscar el nombre completo en el Control de Autoridades y pulse "Lookup". Si el nombre que busca está registrado, podrá incluirlo completo, directamente, seleccionándolo y pulsando "Aceptar", sin necesidad de rellenar la casilla del nombre.

Inicial(es) del nombre con punto(s) y sin espacio entre ellas + nombre entre paréntesis, p. ej. *J.Á. (José Ángel)*

Introduzca su nombre.

**Afiliación: \***

Introduzca su afiliación.

**Idioma: \***

Inglés

Seleccione el idioma de su revisión.

**Derechos de autor, uso y reproducción: \***

Attribution 4.0 International

Derechos de autor, uso y reproducción de su revisión.

**Enlace a los términos sobre el uso y reproducción: \***

<http://creativecommons.org/licenses/by/4.0/>

Enlace a los términos sobre el uso y reproducción de su revisión.

**Tipo: \***

Revisión

Tipo de recurso.

**Importancia del trabajo para su campo: \***

No aplicable

Valore en una escala entre 0-100 la importancia del trabajo para su campo.

**Interés general del trabajo para otros ámbitos académicos: \***

No aplicable

Valore en una escala entre 0-100 el interés general del trabajo para otros ámbitos académicos.

**Importancia del trabajo para la sociedad en general: \***

No aplicable

Valore en una escala entre 0-100 el interés general del trabajo para la sociedad en general (valor social: cuánto podría ayudar esta investigación en los problemas que enfrenta la sociedad actualmente).

**Texto de revisión: \***

Escriba su opinión sobre el trabajo incluyendo todos los detalles necesarios para ayudar a los autores a mejorar su contribución. Procure ser constructivo. Si necesita enviar un texto formateado con figuras, ecuaciones, etc., tendrá la oportunidad a continuación de adjuntar uno o más ficheros adicionales.

**Estándar científico: \***

1

Elija la opción 1 si considera que el trabajo ya es aceptable científicamente. Elija la opción 0 si considera que el trabajo necesita que los autores efectúen una revisión teniendo en cuenta la evaluación. El estándar científico se refiere a varios parámetros relevantes como la metodología aplicada, la claridad de presentación, el uso del idioma, la inclusión de referencias importantes, el rigor metodológico, etc.

**Valoración global del trabajo: \***

0

Valore en una escala entre 0-100 la calidad del trabajo globalmente.

**Subir fichero(s)**

Fichero: \*

Por favor, introduzca la ruta completa del fichero en su ordenador que corresponda con el ítem. Si pincha en "Examinar...", se abrirá una ventana que le permitirá seleccionar un fichero de su ordenador.

**Descripción del fichero: \***

Si lo desea, puede describir brevemente el contenido de este fichero: "Artículo principal" o "Lectura de los datos del documento".

Figura 1. Formulario de evaluación que el revisor recibe por correo electrónico.

Figura 2. El OPRM en repositorios con distintas interfaces y versiones de software.

# Reputación del trabajo

una agregación de sus revisiones,  
**ponderada** por la reputación del revisor



<http://digital.csic.es/handle/10261/130842>

*Figura 3. La reputación del trabajo es una agregación ponderada de las calificaciones globales de sus revisiones, donde el peso es la reputación como revisor del investigador que la realizó.*

# Reputación como autor

una agregación de las reputaciones de sus trabajos,  
**ponderada** por el número de autores



<http://digital.csic.es/handle/10261/130842>

*Figura 4. La reputación como autor es una agregación ponderada de las reputaciones de sus trabajos, donde el peso es el número de autores del trabajo.*

# Reputación de la revisión

una agregación de sus comentarios,  
**ponderada** por la reputación de quien comenta



<http://digital.csic.es/handle/10261/130842>

*Figura 5. La reputación de la revisión es una agregación ponderada de las calificaciones recibidas en los comentarios, donde el peso es la reputación como revisor del investigador que escribió el comentario.*

# Reputación como revisor

una agregación de los comentarios sobre sus revisiones,  
**ponderada** por la reputación del revisor



<http://digital.csic.es/handle/10261/130842>

*Figura 6. La reputación como revisor es una agregación ponderada de los comentarios sobre sus revisiones de los otros revisores, donde el peso es la reputación de los revisores.*



The screenshot displays the e-IEO repository interface. The main content area shows a review page for the paper "Cold-water corals in the Cap de Creus canyon, northwestern Mediterranean: spatial distribution, density and anthropogenic impact". The review is by Alberto Serrano, dated 2016-04-13, with a global reputation score of 80. The interface includes a search bar, navigation menus, and a sidebar with information about the Open Peer Review Module (OPRM).

<http://www.repositorio.ieo.es/e-ieo/handle/10508/10100>    <https://digital.csic.es/handle/10261/131213>

**Figura 7. Las revisiones y los comentarios tienen sus propias colecciones en OPRM.**

## Cold-water corals in the Cap de Creus canyon, northwestern Mediterranean: spatial distribution, density and anthropogenic impact


The screenshot displays the main page for the paper "Cold-water corals in the Cap de Creus canyon, northwestern Mediterranean: spatial distribution, density and anthropogenic impact". The page includes metadata such as the author list (Orejas, C., Gori, A., Lo-Iacono, C., Puig, P., Gili, J.M., Dale, M.R.T.), the publication date (2009), and a global reputation score of 85. The page also features a "Revisión relacionada" section with a score of 80 and a "Cited 76 times in Scopus" badge.

<http://www.repositorio.ieo.es/e-ieo/handle/10508/7818>

**Figura 8. Página de un trabajo en e-IEO con sus métricas.**






### Orejas, C. (Covadonga)



**Biografía**  
Covadonga Orejas is graduate in Biology in Natural Sciences from the University of La Laguna (currently at COB). She has been studying corals and other benthic organisms since 1998 and has been working on the sustainable management of the CWC projects conducted in 12 different geographical areas (44 SCI), and 7 book chapters. She has published in scientific journals, and several funding agencies.

[Descargar en PDF](#)  
[Descargar en fragmento CVN](#)

**Información de contacto**


  

**Campos de especialización**  
Biology and ecology of cold-water corals (CWC) and other benthic organisms


**Grados**  
Graduate in Biology from the Universidad Complutense de Madrid (Spain) (Germany)

**Departamentos**  
C.O. BALEARES

Última actualización marzo 15, 2016

**Reputación del autor**  
 79




### Jerez, S. (Salvador)



**Biografía**  
Salvador Jerez, born in La Gomera (Spain). Researcher at the Island Oceanographic Centre since 1998 with 20 years participating in 19 Spanish projects included in JCR/SC, and 40 National projects including greater amberjack in the European Project.

[Descargar en PDF](#)  
[Descargar en fragmento CVN](#)

**Información de contacto**


  

**Campos de especialización**  
Expert on Greater amberjack (*Seriola dumerili*) aquaculture specializing in

**Grados**  
Graduate and PhD in Biology from the University of La Laguna (Spain)

**Departamentos**  
C.O. CANARIAS

Última actualización marzo 15, 2016

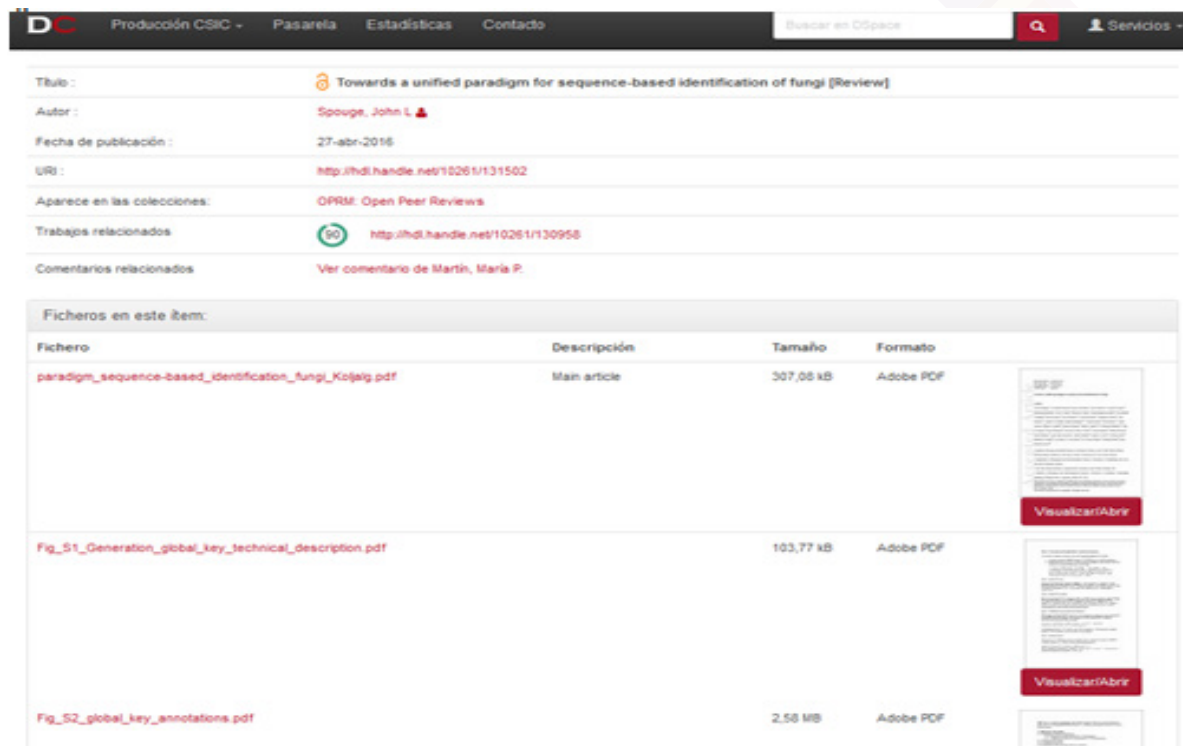
**Reputación como revisor**  
 86



[Ocultar detalles](#)


Mostrando ítems 1-10 de 94


<http://www.repositorio.ieo.es/e-ieo/browse-by>

Figura 9. Reputaciones como autor y como revisor en el perfil de autor de e-IEO.



**DC** Producción CSiC - Pasarela Estadísticas Contacto    Servicios -


**Título:**  Towards a unified paradigm for sequence-based identification of fungi [Review]

**Autor:** Spouge, John L. 

**Fecha de publicación:** 27-abr-2016

**URI:** <http://hdl.handle.net/10261/131502>

**Aparece en las colecciones:** OPRM: Open Peer Reviews

**Trabajos relacionados:**  <http://hdl.handle.net/10261/130958>

**Comentarios relacionados:** [Ver comentario de Martín, María P.](#)

**Ficheros en este ítem:**

Fichero	Descripción	Tamaño	Formato
<a href="#">paradigm_sequence-based_identification_fungi_Kotjaig.pdf</a>	Main article	307,06 kB	Adobe PDF
<a href="#">Fig_S1_Generation_global_key_technical_description.pdf</a>		103,77 kB	Adobe PDF
<a href="#">Fig_S2_global_key_annotations.pdf</a>		2,56 MB	Adobe PDF

<https://digital.csic.es/handle/10261/131502>

Figura 10. La revisión se compone de diversos adjuntos que valoran el trabajo de la investigadora CSiC y el material complementario del artículo. El registro bibliográfico muestra la métrica de reputación ponderada del artículo. La revisión también enlaza con el comentario a la revisión que hace la autora del trabajo original.

English español

Por favor, use este identificador para citar o enlazar a este ítem: <http://hdl.handle.net/10261/133346>

Compartir / Impacto:

EndNote BASE f in t g e

Estadísticas

Título :	 Towards a unified paradigm for sequence-based identification of fungi [Annotation]
Autor :	Martín, María P. 
Fecha de publicación :	11-jun-2016
URI :	<a href="http://hdl.handle.net/10261/133346">http://hdl.handle.net/10261/133346</a>
Aparece en las colecciones:	OPRM: Open Comments
Revisiones relacionadas	 50 Ver revisión de Spouge, John L

Ficheros en este ítem:

Fichero	Tamaño	Formato	
Comment_to_comments.JohnSpouge.docx	36,35 kB	Microsoft Word XML	<a href="#">Visualizar/Abrir</a>

[Mostrar el registro completo](#)<https://digital.csic.es/handle/10261/133346>

*Figura 11. Comentario realizado por la autora de un trabajo sobre la revisión recibida. El comentario, al igual que las revisiones, se muestra como metadato, como fichero adjunto y evalúa también, cuantitativamente, la revisión.*

# “Programa de difusión de la producción intelectual del Instituto Tecnológico de Costa Rica.”

•Autores: *Alexa Ramírez-Vega*

---

**Palabras clave:** visibilidad, contenido académico, acceso abierto, difusión de investigación.

**Keywords:** visibility, academic content, open access, research visibility.

## Introducción

Este documento tiene el objetivo de presentar la propuesta integral para la creación del programa de visibilidad académica del Instituto Tecnológico de Costa Rica (TEC).

A continuación se expone un apartado con los antecedentes referentes a las acciones sobre visibilidad académica emprendidas en el TEC en los últimos años, así como la problemática a la que se enfrenta la institución en cuanto al tema de visibilidad e indicadores de investigación y extensión. Seguidamente se detalla la propuesta de la creación del programa de visibilidad académica, donde se describe su razón de ser y funciones específicas. Además, se presenta un sistema integral para la gestión de la visibilidad del TEC. Finalmente, se exponen los requerimientos, en términos de recursos, para el funcionamiento del programa.

En adelante se deberá entender los siguientes conceptos aquí tratados de la forma que a continuación se describen:

- Presencia web. contar con un espacio o sitio en la web donde se aloja información de una entidad determinada.
- Recuperación de información. recuperar documentos o páginas por terceros a través de las búsquedas en la web.
- Visibilidad web. facilidad con que usuarios de internet recuperan información de terceros. La información debe ser accesible y fácilmente utilizada por los usuarios.
- Acceso abierto. disposición de documentos a texto completo en la web de forma libre para ser accesibles y utilizados por terceros. El acceso se realiza por vía verde (repositorios institucionales) y vía dorada (revistas académicas).
- Generación de indicadores de investigación y extensión. Estos son publicaciones indexadas en web of science, scopus, etc etc.

## Antecedentes

El acceso a la producción científica es uno de los indicadores clave en el momento de evaluar el grado de excelencia y desarrollo de la generación y divulgación de conocimiento de una institución o país determinado. Aquí la visibilidad juega un papel predominante, haciendo más fácil el acceso a los contenidos producidos en las instituciones educativas.

El uso de Internet, y especialmente de la Web 2.0, como medio principal de intercambio de información ha hecho que cualquier persona pueda tener acceso a la producción científica proveniente de instituciones educativas, centro de investigación y academias en todo el mundo. Esta producción es uno de los indicadores clave en el momento de evaluar el grado de excelencia y desarrollo de la generación de conocimiento de una institución o país determinado (Machín, Lugo, & Díaz, 2013). Por lo tanto, la divulgación de los resultados de investigación adquiere mayor relevancia para los investigadores, quienes además de concluir satisfactoriamente los proyectos de investigación, deben dar a conocer los aspectos más relevantes de su labor y así contribuir al mejoramiento y aumento de la producción científica de su institución.

Por su parte, para medir la producción científica se cuenta con variados indicadores, que van desde la incorporación en índices y bases de datos de revistas, rankings mundiales de universidades, hasta la presencia en la web. En todos ellos la visibilidad juega un papel predominante, de manera que la inclusión de las publicaciones periódicas en índices y bases de datos reconocidas, además de prestigio, brinda mayor visibilidad a los documentos que ahí se publican; de igual forma los rankings de universidades como el Webometrics Rankings of World Universities del Laboratorio de Cibermetría del Consejo Superior de Investigación Científica de España (CSIC, España, 2014) que es calculado basado principalmente en el contenido presente en la web, proveniente de diferentes bases de datos académicas; finalmente, la presencia en la web dependerá de la cantidad de contenidos disponibles en Internet y la posibilidad de ser recuperados por buscadores especializados.

Esto ha impulsado a las instituciones académicas a establecer estrategias y programas dedicados a la divulgación de sus contenidos en la web así como al seguimiento y evaluación del impacto de esta divulgación en los indicadores obtenidos de diversos sitios como ResearchGate, Google Scholar citations, rankings web universitarios, etc.

De esta manera, la visibilidad de los contenidos en la web deben tener dos características principales: disponibilidad y confiabilidad. La disponibilidad en cuanto el contenido pueda ser recuperado por cualquier lector en la web; y confiabilidad en cuanto el contenido recuperado sea catalogado como académico y se logre identificar su autor y procedencia.

En el año 2014 el TEC realizó una valoración de los contenidos producidos por el personal del Instituto Tecnológico de Costa Rica (TEC) y su disponibilidad en la web. Así se evaluaron aspectos como la cantidad de documentos colocados en la web, tipos de documentos y la disponibilidad de estos contenidos.

Algunos de los problemas de visibilidad detectados fueron los siguientes:

- Dos dominios web principales para la institución (itcr.ac.cr y tec.ac.cr). Esto divide la producción y páginas del TEC en dos dominios que se podrían interpretar como dos entes distintos y afectar la visibilidad del TEC como institución. Se inició con la unificación de los dominios hacia tec.ac.cr a partir de octubre del 2014. Y para mayo del 2015 ya se contaba con el 90% de los contenidos migrados.
- Mal posicionamiento del TEC en ranking de visibilidad web: para el 2014 el TEC se encontraba en la posición 10 a nivel nacional y en la posición 6297 a nivel mundial.
- Los artículos publicados en revistas del TEC no estaban disponibles en Internet y no se encontraban en un sitio centralizado. Para esto se creó el portal de revistas académicas a partir del 2012. Para 2015 se cuenta con siete revistas alojadas en este sistema.
- Los trabajos publicados por los investigadores, en otras revistas o sitios, no se encuentran disponibles para su consulta en una base de datos centralizada. Además de la ficha del investigador no es accesible por terceros, para obtener información adicional sobre su producción, con quién y dónde la publican.
- Las memorias de congresos de los eventos realizados en el TEC no se encuentran en Internet. Esto dificulta el acceso a este contenido y la divulgación de los resultados expuestos en estos congresos y conferencias. En septiembre de 2015 se creó un espacio para publicar memorias de congresos organizados por alguna entidad del TEC. Se cuentan con 4 memorias colgadas en el sitio. Falta un seguimiento sistemático para la recuperación y depósito de nuevas memorias editadas por el TEC.
- De los documentos localizados en la web se encontró inconsistencias de los nombres de los autores y la institución. Los nombres de los autores se encontraron poco o nada normalizados, lo cual afecta la forma de citar correctamente los documentos e identificar la autoría correcta.

Dado lo anterior, el TEC inició algunas acciones alistados para aumentar la visibilidad de sus resultados de investigación a través de la vía verde y la vía dorada del Acceso Abierto (Hernández Pérez, Rodríguez Mateos, & Bueno De la Fuente, 2007), con el fin de aumentar la cantidad de contenidos en la web, garantizar su disponibilidad y mejor la confiabilidad de éstos. Aunque estos esfuerzos han generado algunos resultados positivos, no son suficientes, por lo tanto, se ve en la necesidad de un proceso sistemático que permita guiar las acciones de manera integral y sostenida para la mejora continua de la visibilidad académica del TEC.

A pesar, que en el TEC existen diversas unidades especializadas en áreas de gestión y divulgación de la información académica, no se cuenta con una unidad encargada de dirigir y coordinar los diversos procesos para que la información académico-científica del TEC sea visible en la web de acuerdo a las exigencias en el área de visibilidad académica a nivel mundial. De esta manea, se propone la creación de un programa especializado en visibilidad web académica para el TEC.

## Resultados preliminares

Algunos de los problemas de visibilidad detectados se han tratado de corregir gracias a las iniciativas aisladas de diversas dependencias del TEC. Algunas de éstas se describen a continuación:

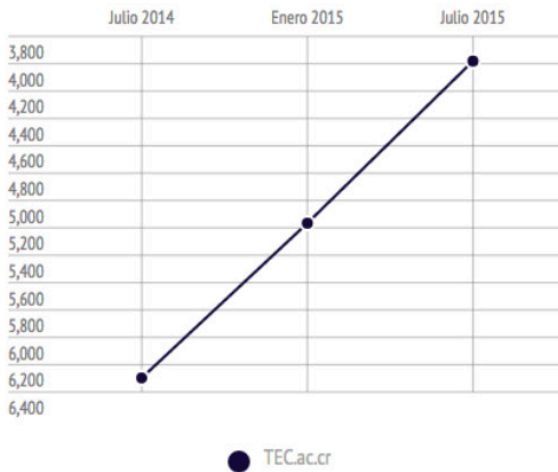
- Unificación de dominios. Por medio de varias reuniones de coordinación con 27 funcionarios del TEC relacionados con sistemas informáticos alojados en los diferentes dominios utilizados por el TEC, donde se valoraron los contenidos alojados en los dominios web, donde se tomó la decisión de trasladar todos los contenidos al dominio único
- tec.ac.cr, esto apoyado por un acuerdo de la Rectoría. De esta manera se arracó con un proceso de migración que dio inicio el 30 de octubre del 2014, donde se estableció un cronograma de migración con miras a finalizar en julio del 2015.
- Mejoramiento de revistas del TEC. Desde 2012 se inició con el mejoramiento de las revistas académicas del TEC, logrando la creación del Portal de revistas; inclusión en diversos índices y bases de datos; capacitación a editores; entre otras acciones.
- Fortalecimiento de repositorio institucional. Aunque se creó en 2004 como un repositorio de la biblioteca, en 2013 se formó un grupo para mejorarlo y convertido en repositorio institucional. Para esto se actualizó el sistema a la versión mas reciente, se reestructuró la distribución de los contenidos, también se redactaron políticas de repositorio donde se consigna la razón de ser del sitio, así como las normas para el deposito de material.
- Creación de Portal de Memorias de Congresos. En octubre de 2015 se lanza el sitio Memorias TEC, el cual tiene el objetivo de publicar las memorias de actividades académicas (congresos, conferencias, simposios, etc.) organizadas por alguna dependencia del Instituto Tecnológico de Costa Rica. Esto con el fin de poner a disposición de la comunidad nacional e internacional, por medio de Internet, los trabajos presentados en dichos eventos. De esta manera se contribuye con la presencia y visibilidad de la producción científica que se genera como parte de las actividades del TEC.
- Capacitación en uso de Google Scholar. Desde 2013 se inició con la capacitación a investigadores en el uso de Google Scholar, haciendo énfasis en la creación del perfil de citas, de forma que cada participante tome consciencia de la importancia de esta herramienta, así como pautas para localizar sus publicaciones en la web. El taller fue impartido en 10 fechas distintas, en varias sedes del TEC, donde se logró capacitar a más de 150 investigadores, quienes crearon su perfil en Google Scholar y lograron incorporar su producción científica disponible en Internet. Además, se logró detectar problemas con la normalización de los nombres de los autores, así como el nombre de la institución, el cual aparecía de hasta cinco formas distintas en la web.
- Divulgación. En diversas instancias del TEC se han iniciado esfuerzos que permitan divulgar información de relevancia del quehacer del TEC, a manera de noticias digitales. Los medios que se han creado son:
  - » Hoy en el TEC.
  - » Revista Pensis.
  - » Portal de noticias Investiga TEC.
- Mejora de web institucional. Se creó un grupo de trabajo encargado de reestructurar la web principal del TEC. Esto ha permitido completar una primer etapa donde se implementó la página de inicio del web, basado en diversos estudios a usuarios y una investigación sobre arquitectura de información, la cual permitió fundamentar en



nuevo diseño del web, actual, dinámico y adaptado a dispositivos móviles.

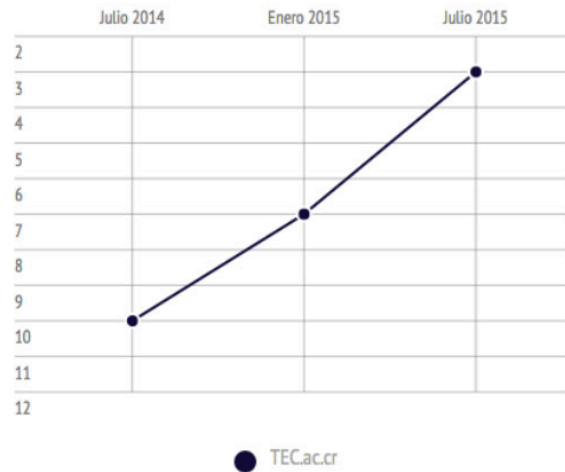
Uno de los principales medios para medir el impacto de las acciones anteriores, en términos de visibilidad académica, es el resultado obtenido en el ranking Webometrics, el cual ha permitido constatar el avance del TEC en visibilidad web. Se pasó de la posición 6200 a nivel mundial a la posición 3981 en menos de un año; a nivel nacional se pasó de la posición 10 a la 4 (como se muestra en la figura 2).

## En el mundo



Posición del TEC en el ranking mundial Webometrics, que analiza los portales en Internet de unas 12.000 casas de enseñanza.

## En Costa Rica



Comparación con otras universidades costarricenses que también son analizadas en Webometrics.

**Figura 2. Avance en posición nacional e internacional en ranking webometrics**

Finalmente, la Web es clave para el futuro de todas las misiones de la universidad, puesto que es de hecho la principal herramienta de comunicación académica, el canal primordial para desarrollar la educación a distancia, un espacio para la participación de la comunidad y el escaparate abierto al mundo para atraer talento, financiación y recursos de todo tipo.

## Propuesta inicial

Creación de un programa dentro de la VIE encargada de dirigir y coordinar con los diferentes entes institucionales los procesos de sistematización de los contenidos académicos con el objetivo de lograr una mayor visibilidad en la web de las capacidades y producción del TEC. Además del monitoreo y evaluación de indicadores en diferentes plataformas.

## Misión

Dar mayor visibilidad y accesibilidad en la web a las capacidades y producción científica e intelectual de las distintas unidades académicas del TEC. Así como evaluar los distintos indicadores web sobre visibilidad de estos contenidos.

## **Visión**

Ser una unidad especializada en visibilidad a nivel nacional e internacional que facilite y promueva la divulgación de las capacidades y producción intelectual del TEC en la web.

## **FUNCIONES**

La unidad de visibilidad se encarga de dirigir, armonizar y coordinar los diferentes procesos y funciones que se muestran en la figura 1. Para ello deberá cumplir las siguientes funciones específicas:

- Coordinar con las diferentes instancias institucionales encargadas de la gestión y divulgación de la información académica, para garantizar la visibilidad de los contenidos colocados en la web.
- Participar en los procesos de creación de políticas institucionales orientadas a mejorar la visibilidad de los contenidos académicos en la web.
- Desarrollar herramientas que faciliten la colocación de contenidos en la web para lograr mayor visibilidad de los contenidos académicos.
- Definir las pautas requeridas para guiar las diversas acciones institucionales relacionadas con la mejora de la visibilidad en la web.
- Evaluar las acciones y los procesos desarrollados por la institución en materia de visibilidad, mediante indicadores establecidos.
- Promover actividades de capacitación y asesoría en materia de visibilidad dirigida a investigadores, profesores y funcionarios del TEC.
- Desarrollar investigaciones sobre temas relacionados con visibilidad institucional.
- Definición y estimación de indicadores para medir la efectividad de las acciones realizadas para mejorar la visibilidad.
- Apoyar los procesos institucionales sobre presencia en rankings nacionales e internacionales.
- Desarrollar estrategias de divulgación y mercadeo digital para posicionamiento de los contenidos académicos en la web.
- Crear un sistema integral de visibilidad institucional y orientar las acciones concretas para su implementación.

## **SISTEMA INTEGRAL DE VISIBILIDAD DEL TEC**

En la figura 1 se muestra el sistema integral de visibilidad académica propuesto para el TEC. Este sistema se compone de una serie de elementos agrupados según su función y objetivo, los cuales se describen a continuación.

## Propuesta Visibilidad Académica TEC



Figura 1. Sistema integral de visibilidad institucional del TEC

### PUBLICACIÓN

Repositorio institucional. Lugar donde se deposita toda la producción científica y académica producida por el TEC. Aquí se incluyen: informes de investigación, tesis y trabajos finales, material didáctico, material multimedia, pre y post prints de artículos, artículos publicados en revistas del TEC, memorias de congresos organizados por entidades del TEC y ponencias de funcionarios del TEC presentados en congresos externos.

- Portal de revistas. Sitio especializado y centralizado desarrollado en OJS (Open Journal Systems) donde se alojan todas las revistas editadas en el TEC.
- Memorias de congresos. Sitio centralizado donde se alojarán todos los datos principales de los eventos académicos y los documentos oficiales (memorias) productos de las actividades desarrolladas en los congresos organizados por alguna entidad del TEC.

### NORMATIVA

- Política de repositorio. Normas que permiten regular y definir pautas a seguir en cuanto al depósito y uso del repositorio Institucional del TEC.
- Política de acceso abierto. Política de acceso abierto para el TEC, donde la institución se pronuncie en este tema, además se defina las acciones que realiza para su apoyo y fortalecimiento.
- Normativa de visibilidad para investigadores. Creación de perfiles en redes académicas, divulgación de publicaciones en la web, etc.

## **CAPACITACIÓN Y ASESORÍA**

Programas para capacitar y asesorar a las instancias interesadas (profesores, investigadores, etc.) en temas que les permita mejorar la visibilidad y calidad de sus publicaciones.

## **DIVULGACIÓN**

- Portal Investiga TEC. Sitio donde se aloja información relevante para investigadores. Entre las cuales se encuentran:
  - » Contenido original relacionado con investigación y extensión sobre el quehacer del TEC.
  - » Boletín electrónico mensual enviado a los suscriptores vía correo electrónico con la información más relevante publicada en el portal.
  - » Divulgación estratégica mediante redes sociales de las actividades sobre investigación y extensión del TEC.
  - » Información relevante (noticias, actividades, etc.) relacionadas con investigación y extensión.
- Hoy en el TEC. Boletín electrónico institucional encargado de divulgar noticias del quehacer del TEC, a cargo de la Oficina de Comunicación y Mercadeo.
- Pensis. Revista divulgativa en formato electrónico sobre temas de actualidad en el campo académico. A cargo de la Oficina de Comunicación y Mercadeo.
- Página web. Página web principal del TEC. A cargo de la Oficina de Comunicación y Mercadeo.

## **INVESTIGACIÓN**

Realizar investigaciones que permitan detectar debilidades y oportunidades de mejora de la producción, visibilidad y publicación de trabajos producidos por investigadores del TEC.

Algunas líneas de investigación:

- Métricas alternas para medir el impacto de las publicaciones del TEC.
- Estudio sobre la falta de normalización de los nombres de autores en artículos científicos.
- Indicadores de visibilidad de la producción científica del TEC.
- Índice H de investigadores del TEC según Google Scholar y ResearchGate.
- Evaluar portales del TEC mediante Search Engine Optimization (SEO).
- Estudios sobre rankings internacionales.
- Evaluación y monitores de impacto en indicadores, como avanzamos, que cambios de tendencias hay y que sectores del TEC avanzan en materia de visibilidad.

## **Conclusiones**

Esta propuesta se ha trabajado de manera aislada por diversas dependencias de la institución, la idea es unificar los esfuerzos en un programa que se encargue de todas estas acciones, con el objetivo de realizar actividades sistematizadas y constantes que mejoren la visibilidad de la producción científica de la institución.

## Referencias

CSIC, España. (2014). Webometrics Rankings of World Universities. Retrieved May 16, 2014, from <http://www.webometrics.info/en/Methodology>

Hernández Pérez, T., Rodríguez Mateos, D., & Bueno De la Fuente, G. (2007). Open Access: el papel de las bibliotecas en los repositorios institucionales de acceso abierto. Retrieved from <http://digitum.um.es/xmlui/handle/10201/4017>

Machín, M. P., Lugo, Y. G., & Díaz, M. del P. F. (2013). Análisis bibliométrico sobre la visibilidad científica y el factor de impacto en ciencias de la educación. *EDUMECENTRO*, 3(3), 99–112.

# “Consolidación de la Red Mexicana de Repositorios Institucionales – REMERI e iniciativas de desarrollo y colaboración”

•Autores: Rosalina Vázquez Tapia. Antonio Felipe Razo Rodríguez.

---

## Resumen

Este trabajo muestra el estado actual de la Red Mexicana de Repositorios Institucionales REMERI y las actividades realizadas para su desarrollo y consolidación como la red nacional representando a México en el proyecto LA Referencia. Se explican las actividades de difusión y trabajo con la comunidad de repositorios en México y la promoción en eventos internacionales y redes sociales. Se presentan los servicios desarrollados con la plataforma INDIXE y el esquema de metadatos para interoperabilidad con LA Referencia y OpenAIRE. También se muestran las actividades de realizadas en el marco del Repositorio Nacional de Ciencia y Tecnología de CONACyT. Por último, se explica la colaboración con la iniciativa de la Enciclopedia de la Literatura en México y sus resultados. This work shows the current state of the Mexican Network of Institutional Repositories REMERI and activities for development and consolidation as the national network representing Mexico in the project LA Referencia. We present the dissemination activities and community work in Mexico to promote REMERI in international events and social networks. We present services developed with INDIXE platform and metadata schema for interoperability with the reference and OpenAIRE. We show the activities related to de Science and Technology National Repository from CONACyT. Finally, we explain the collaboration with the initiative of the Encyclopedia of Literature in Mexico and their results.

**Palabras clave:** Repositorios Institucionales, Acceso Abierto, Vía Verde, Vía Dorada, Archivos, Interoperabilidad

## Introducción

La Red Mexicana de Repositorios Institucionales – REMERI <http://www.remeri.org.mx> (Figura 1), fue creada en 2012 como una iniciativa de 6 Instituciones de Educación Superior bajo la coordinación de la Universidad Autónoma de San Luis Potosí (UASLP) y con financiamiento del Consejo Nacional de Ciencia y Tecnología CONACyT de México. El objetivo de REMERI es integrar, difundir y dar visibilidad a la producción científica, académica y documental de las instituciones miembros e incorporarse a redes o agregadores de repositorios internacionales para fomentar la colaboración y la divulgación de contenidos de Acceso Abierto en español. Actualmente (agosto de 2016) REMERI integra más de 517,000 registros de 124 Repositorios Institucionales de 71 instituciones



mexicanas (Figura 2) y desde el 2012 es el nodo nacional que representa la participación de México en la Red Federada de Repositorios Institucionales de Publicaciones Científicas LA-Referencia (<http://www.lareferencia.info>), aportando la mayor cantidad de recursos en el idioma español en este proyecto con 128,604 registros en agosto de 2016. Recién fue aprobado REMERI como proveedor de datos en OpenAIRE aportando 119,352 documentos en acceso abierto en agosto 2016.



**Figura 1. Red Mexicana de Repositorios Institucionales**

Desde el 2013 REMERI es una comunidad académica de la Corporación Universitaria para el Desarrollo de Internet CUDI (<http://www.cudi.edu.mx>) que cuenta con más de 350 miembros, organizando semestralmente reuniones, mesas de conferencias y cursos de capacitación especializados; además, para la visibilidad internacional de la Red, participa con ponencias y publicaciones en eventos internacionales. Se tuvo participación en Open Repositories 2015 y 2016 uno de los eventos en el área más relevantes del área a nivel internacional y en los encuentros latinoamericanos de TICAL 2014 y 2015. Para la promoción a través de las redes sociales, se cuenta ya con una página de Facebook y una cuenta de Twitter, además de la lista de correo de la comunidad REMERI-CUDI. En el contexto de las reuniones semestrales de CUDI, se han organizado tres mesas de conferencias sobre temas de gestión de derechos de autor en el entorno digital, interoperabilidad y servicios especializados, además de dos talleres, uno sobre la gestión y administración del software Dspace y el otro sobre los fundamentos técnicos para la interoperabilidad de los repositorios institucionales con OpenAIRE y el Repositorio Nacional.

DICIEMBRE 2012  
FUNDADORES



Figura 2. Instituciones participantes en REMERI

## Desarrollo y resultados

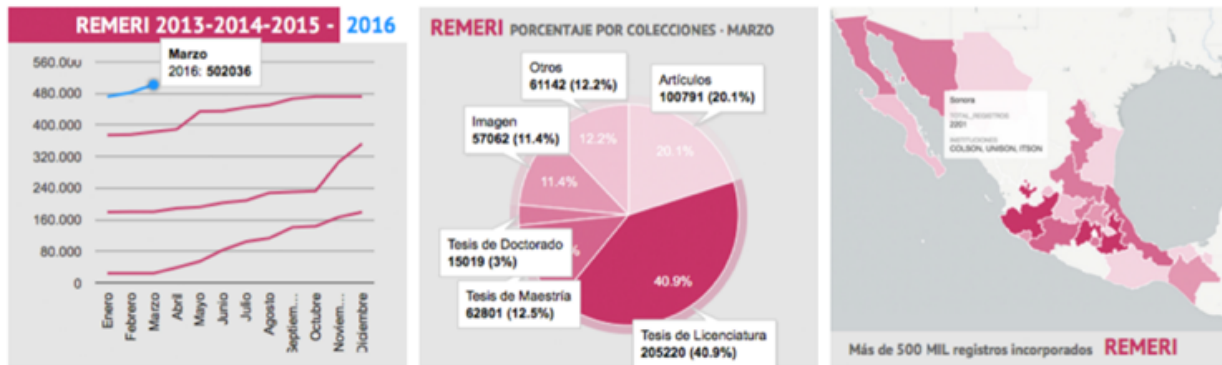
REMERI desarrolló una plataforma tecnológica denominada INDIXE de acuerdo a los requerimientos y lineamientos técnicos junto con las características de interoperabilidad de los Repositorios Institucionales en México y las necesidades del proyecto. INDIXE está desarrollado en el lenguaje xQuery y trabaja con una base de datos XML. Cuenta con servicios de recolección de metadatos, normalización e integración de registros, sistema de búsqueda y consulta por relevancia y por filtros, así como servicios de interoperabilidad como servidor de metadatos OAI-PMH. El portal y los servicios se encuentran desarrollados en JSP y Java en un servidor Tomcat con la base de datos eXist. Cuenta con su propio servidor en el centro de datos de la Universidad Autónoma de San Luis Potosí y un servidor de respaldo proporcionado por CUDI.

Utilizando la plataforma INDIXE fue posible replicar la metodología de trabajo y la solución a otras iniciativas y servicios [1], de tal manera que se pueden consultar los siguientes servicios (agosto 2016):

- El INDIXE de Tesis Digitales de instituciones mexicanas con más de 270 mil tesis de licenciatura, maestría y doctorado de 53 universidades y centros de investigación: <http://www.remeri.org.mx/tesis>
- El INDIXE de Revistas y Publicaciones Periódicas de Acceso Abierto de instituciones mexicanas con 350 revistas y más de 100 mil artículos: <http://www.remeri.org.mx/revistas>
- El INDIXE de Archivos y Fondo Antiguo (en desarrollo) con más de 12,000 materiales de 7 archivos y bibliotecas: <http://www.remeri.org.mx/archivos>

Además, REMERI cuenta con un directorio de todos los Repositorios Institucionales que integran la red, con su información detallada <http://www.remeri.org.mx/repositorios> y un

indicador de producción científica institucional de acuerdo a los materiales recolectados y está disponible en la siguiente dirección: <http://www.remeri.org.mx/producción> . Todos los servicios de consulta cuentan con versiones responsivas para dispositivos móviles. También se desarrolló una versión del portal y del buscador en inglés para promover la internacionalización del proyecto. El portal además incluye nuevas estadísticas y líneas de tiempo y mapas interactivos (Figura 3).



**Figura 3. Estadísticas y mapas interactivos**

REMERI participa actualmente en la segunda etapa del proyecto LA Referencia siguiendo los lineamientos del documento de “Metadatos y Políticas de Cosecha de LA Referencia” con un cumplimiento del 80% de los lineamientos y especificaciones técnicas (Figura 4).

REMERI a participado desde el 2012 con pruebas de conectividad y de forma continua desde la primera versión en el 2013 y con la incorporación continua en registros cosechados como se muestra en el figura 5. : 3,500 (2012) – 68,507 (2013) – 101,979 (2014) – 120,698 (2015) - 128,604 (2016).

REMERI apoya de manera permanente a las instituciones mexicanas en el desarrollo y consolidación de sus repositorios digitales. Promueve y facilita la normalización de metadatos, la adopción de estándares de interoperabilidad y el cumplimiento de los lineamientos técnicos del Repositorio Nacional en asesorías y talleres.

DUBLIN CORE	REMERI CORE	DUBLIN CORE	LA-REFERENCIA
dc:title	titulo	dc:title	dc:title
dc:description	descripción	dc:description	dc:description
dc:language	idioma	dc:language	dc:language
dc:creator	autor (múltiple)	dc:creator	dc:creator
dc:date	fecha	dc:date	dc:date
dc:type	tipo	dc:type	dc:type
dc:identifier	link	dc:identifier	dc:identifier
dc:subject	tema (múltiple)	dc:subject	dc:subject
dc:source	fuentes	dc:source	dc:source
dc:relation	referencia (múltiple)	dc:relation	dc:relation
dc:coverage	cobertura (múltiple)	dc:coverage	dc:coverage
dc:publisher	editor	dc:publisher	dc:publisher
dc:contributor	colaborador	dc:contributor	dc:contributor
dc:format	formato	dc:format	dc:format
dc:rights	derechos	dc:rights	dc:rights
	institución	dc:identifier	dc:identifier>instname:
	siglas (de la institución)	dc:identifier	dc:identifier>instname:
	repositorio	dc:identifier	dc:identifier>reponame:
	url repositorio	dc:identifier	dc:identifier>repourl:

**Figura 4. Esquema de Metadatos Dublin Core, REMERI y LA Referencia**

Durante el primer cuatrimestre de 2016 REMERI colaboró con el proyecto de la Enciclopedia de la Literatura en México o ELEM <http://www.elem.mx> (Figura 5) que integra la información de más de 10,000 autores y 30,000 obras literarias mexicanas. Tomando como referencia la Biblioteca Virtual Miguel de Cervantes (<http://www.cervantesvirtual.com>) y su plataforma de desarrollo Cervantes Virtual Research (<http://bvmcresearch.cervantesvirtual.com>) donde podemos consultar las diferentes soluciones que han desarrollado para la consulta y estructura de su colección [2].



**Figura 5. Gráfica por año**

Se trabajó en colaboración con la ELEM para localizar en los Repositorios Institucionales obras de autores mexicanos, artículos de investigación y tesis con menciones tanto de obras como de autores mexicanos, así como tesis y revistas de las escuelas de literatura del país, además de promover la normalización e identificación de autores con identificadores en ambas plataformas. Se desarrollaron servicios de interoperabilidad para la ELEM con el protocolo OAI-PMH con base en la experiencia de REMERI, por su parte la codificación semántica en RDF de la información de las obras literaria y la adopción de identificadores para Linked Data se planea adoptar en REMERI.



**Figura 5. Portada de la Enciclopedia de la Literatura en México, ELEM.**



## Conclusiones

En los últimos dos años de trabajo y consolidación de la red, casi se han duplicado el número de repositorios, instituciones y registros incorporados en su plataforma, además de desarrollar nuevos servicios de consulta e interoperabilidad. El INDIXE de Tesis Digitales promueve la Vía Verde del Acceso Abierto, mientras que el INDIXE de Revistas promueve la Vía Dorada. La colaboración con otras iniciativas como la ELEM promueve la estandarización, la interoperabilidad y proporciona un valor agregado a los estudiantes e investigadores. Estos servicios y modelos de colaboración son un referente para replicar iniciativas en otros países latinoamericanos y siguiendo los pasos de LA-Referencia proponemos crear un catálogo latinoamericano de tesis digitales, un servicio integrado de revistas de acceso abierto latinoamericanas y una enciclopedia de la literatura latinoamericana.

En esta propuesta presentamos como ha sido la evolución de la Red hasta su consolidación, los servicios que ofrece, las perspectivas de desarrollo y las áreas de oportunidad de nuevos proyectos colaborativos en el marco internacional. REMERI es una iniciativa reconocida tanto a nivel nacional e internacional como un proyecto innovador y es un referente para LA Referencia representando a CONACyT en su compromiso de proveer un nodo nacional.

## Bibliografía

Candela Romero, G., Escobar Esteban, M. P., Marco Such, M., Pernías Peco, P., & Sánchez Díaz, A. (2012). A Linked-Data based approach for managing digital libraries. Proceedings of the SPDECE-2012. Ninth Multidisciplinary symposium on the design and evaluation of digital content for education.

Vazquez R., Razo A. 2015. El desarrollo de servicios especializados para promover el Acceso Abierto: el INDIXE de Tesis Digitales y el de Revistas. Quinta Conferencia de Directores de Tecnología de Información, TICAL2015. Viña del Mar, Chile. Julio 2015.

Vazquez R., Razo A. 2015. Federated Networks of Open Access Repositories in Mexico and Latin America. Open Repositories 2015. Indianápolis, Indiana USA. Junio 2015.

Vazquez R., Razo A. 2014. INDIXE: el agregador Nacional de la Red Mexicana de Repositorios Institucionales-REMERI para la Interoperabilidad con Redes Federadas de Repositorios Institucionales. Conferência Internacional Acesso Aberto, Preservação Digital, interoperabilidade, Visibilidade de Dados Científicos. Porto Alegre, Brasil. Octubre 2014.

Vazquez R., Razo A. 2014. Interoperabilidad entre Redes Federadas de Repositorios Institucionales para la disseminación del conocimiento: El caso REMERI de México. Cuarta Conferencia de Directores de Tecnología de Información, TICAL2014. Cancún, México. Mayo 2014



# “Estrategia de Visibilidad de Recursos Educativos Abiertos para el Sistema de Educación Costarricense a través de un Repositorio Institucional”

•Autores: *Isaac Alpizar Chacon, Laura Meneses Guillen.*

## Resumen

El material didáctico generado en la universidad es de gran calidad, y con mucho potencial para ser utilizado por un gran número de estudiantes y profesores. Sin embargo, estos elementos educativos por si solos no pueden ser aprovechados por estudiantes de escuelas y colegios, pues ellos no tienen una forma de acceder y de buscar estos recursos educativos. Para solucionar este problema, el Tecnológico de Costa Rica diseño una estrategia de visibilidad de material didáctico y recursos digitales producidos en la universidad, para que los estudiantes puedan acceder a ellos a través de Recursos Educativos Abiertos. La estrategia contempla la creación de Recursos Educativos abiertos, el uso del Repositorio Institucional para dar visibilidad a estos recursos, y la utilización de un estándar de metadatos especializado. Como resultados esperamos contribuir con la educación de estudiantes y profesores, potenciar el repositorio, y dar mayor visibilidad al trabajo de calidad producido en la universidad.

**Palabras clave**—visibilidad, recursos educativos abiertos, dspace, repositorio institucional, LOM.

## Abstract

The educational materials generated at the university have high quality and a huge potential to be useful to a large number of students and teachers. However, these educational resources themselves can not be used by students in schools because they don't have a way to access and search for these educational resources. To solve this problem, the Tecnológico de Costa Rica designed a visibility strategy to transform all this educational content into Open Educational Resources to be used by students. The strategy includes the creation of OER, the use of the institutional repository to give visibility to these resources, and the use of a specialized metadata standard. As results we expect to contribute to the education of students and teachers, strengthen the repository, and give greater visibility to the quality work produced at the university.

**Keywords**—visibility, open educational resources, dspace, institutional repository, LOM.

## 1. Introducción

El material didáctico que genera una universidad usualmente es de gran calidad educativa, abarca una gran variedad de temas, y tiene potencial para ser usado por estudiantes en todos los niveles del sistema educativo, desde primaria hasta la educación superior. Sin embargo, este material usualmente queda destinado únicamente a estudiantes de cursos presenciales en la universidad, sin posibilidad de ser accedido por otros estudiantes que podrían beneficiarse con dichos recursos educativos. Por ejemplo, en el Tecnológico de Costa Rica (TEC) se ha desarrollado una gran cantidad de material didáctico para diferentes cursos (Química I, Química II, Matemática General, Cálculo Diferencial e Integral, Física I, Física II, Circuitos de Corriente Alterna, Circuitos de Corriente Continua, Biología, Introducción a la Programación, y otros) porque se ha trabajado en un proceso de virtualización de los cursos con mayor demanda y nivel de repetición. De este proceso, sale una gran cantidad de recursos (definiciones, ejercicios, imágenes y diagramas) y materiales audiovisuales (videos, animaciones) que además de ser útiles para estudiantes de esos cursos, podría beneficiar a estudiantes de primaria y secundaria. Sin embargo, estos recursos por si solos no pueden ser aprovechados por alumnos de escuelas y colegios, pues ellos no tienen una forma de acceder y de buscar estos recursos educativos.

Ante esta situación, el Tecnológico de Costa Rica diseñó una estrategia de visibilidad de material didáctico y recursos digitales producidos en la universidad, para que los estudiantes puedan acceder a ellos a través de Recursos Educativos Abiertos.

El término “Recursos Educativos Abiertos” o REA fue acuñado en el año 2002 por la UNESCO y son todos los “materiales de enseñanza, aprendizaje, o investigación que se encuentran en el dominio público o que han sido publicados con una licencia de propiedad intelectual que permite su utilización, adaptación y distribución gratuitas” (UNESCO). En el 2012 se realizó el Congreso Mundial de Recursos Educativos Abiertos, donde se emitió la Declaración de París que contempla una serie de recomendaciones a los estados para fomentar y promover el conocimiento a través del uso y la creación de recursos educativos abiertos (Vidal Ledo, Alfonso Sánchez, Zacca González, & Martínez Hernández, 2013). En (Butcher, 2015) se presenta una guía básica de REA y en (Santos-Hermosa, Ferran-Ferrer, & Abadal, 2012) se presenta una revisión bibliográfica de los REA, su disposición en repositorio y su uso.

La estrategia de visibilidad contempla el uso del Repositorio TEC, como repositorio institucional de la universidad para hacer visible todos los REA que se generan y así puedan ser usados por una población más allá de los estudiantes matriculados en la universidad. Además, se contempla la creación y utilización de un estándar de metadatos especializado para etiquetar los REA de una mejor manera. Con esta estrategia se espera: 1) poner a disposición de profesores y estudiantes de primaria y secundaria, nuevos recursos de calidad que les sean útiles en el desarrollo de sus actividades académicas, 2) fortalecer el uso del repositorio institucional o Repositorio TEC al enriquecerlo con contenido de calidad, y 3) dar mayor visibilidad al trabajo de calidad que se produce en la universidad.

El resto del artículo está estructurado de la siguiente manera. En la sección 2 se presenta el detalle de la estrategia de visibilidad desarrollada. Los resultados parciales se comentan en la sección 3 y se hace una discusión de los posibles beneficios de aplicación de

la estrategia de visibilidad. Finalmente en la sección 4 de conclusiones se resumen los principales aportes de este trabajo.

## 2. Materiales y metodología

Teniendo como meta generar Recursos Educativos Abiertos de calidad a partir de los materiales didácticos desarrollados por los profesores del TEC, y con la intención de que estos sean de utilidad para estudiantes de primaria y secundaria de Costa Rica, se desarrolló una estrategia para crear y dar visibilidad a estos recursos. Debido a la importancia de este proyecto, la elaboración y puesta en marcha de dicha estrategia se realizó en conjunto por entes claves. Primero, la Vicerrectoría de Investigación y Extensión (VIE) como enlace entre la universidad y el Gobierno de Costa Rica en temas de investigación y cooperación empezó a articular la idea de poner a disposición del sistema educativo costarricense material didáctico creado en la docencia universitaria. Luego, la Biblioteca José Figueres Ferrer como administradora del Repositorio Institucional, visualizó una clara oportunidad para potenciar el Repositorio TEC, y su capacidad para hacer accesible todos estos recursos. Finalmente, el TEC Digital, como unidad encargada de promover y facilitar los procesos de enseñanza – aprendizaje a través del uso de tecnologías de información y comunicación, jugó un papel clave al aportar el conocimiento técnico y pedagógico para crear los REA. Del trabajo de estos tres departamentos, surge una estrategia institucional para crear y dar visibilidad a Recursos Educativos Abiertos para ser utilizados en el sistema de educación costarricense, a través del Repositorio Institucional como medio para dar acceso a dichos recursos. A continuación se detalla dicha estrategia, primero mencionando los procesos y políticas que abarca, y luego describiendo los detalles técnicos necesarios para su realización.

### 2.1 Definición de procesos y políticas

La estrategia contempla un proceso claro, donde diferentes actores entran en juego para realizar las diferentes tareas necesarias para la ejecución del proyecto. Se mencionan a continuación de manera ordenada, cada una de las tareas contempladas en la estrategia de visibilidad

- I. **Convocatoria inicial:** La VIE realizó una divulgación institucional, invitando a profesores y estudiantes que contaran con material didáctico, aplicaciones y cualquier otro recurso virtual a ponerlo a disposición, para construir los Recursos Educativos Abiertos y ponerlos a disposición del Ministerio de Educación de Costa Rica (MEP). El MEP es el ministerio regulador de la educación pública y privada en los niveles de escuela y de colegio. Esta convocatoria fue posible ya que la universidad firmó un convenio de cooperación con el Gobierno de Costa Rica, por lo que se han podido desarrollar estas iniciativas para compartir y dar mayor visibilidad a los trabajos desarrollados en el TEC.
- II. **Levantamiento de material didáctico:** A partir de la convocatoria inicial, un grupo de profesores mostró su interés de compartir su material didáctico en la iniciativa planteada. En este punto se pudo valorar que había disposición y material suficiente para llevar a cabo el proyecto, y continuar con las siguientes tareas.
- III. **Selección de coordinador:** Debido a que los profesores con recursos digitales provienen de diferentes Escuelas y áreas, se decidió nombrar un coordinador por cada una de las Escuelas, y así formar un grupo de comunicación más reducido. Cada uno

de estos profesores coordinadores reciben la información del proyecto, y luego ellos la pasan con los demás profesores de su Escuela que están participando.

- IV. **Creación y distribución de un formulario para la creación de REA:** Con el propósito de facilitar a los profesores la descripción de cada uno de los materiales didácticos que ponen a disposición para compartirlos como REA, se creó un formulario editable en PDF. Este formulario permite que los profesores puedan llenar valores textuales (por ejemplo, el nombre del RE y los autores), y que puedan seleccionar de forma fácil con un check otros valores (por ejemplo, el nivel de dificultad). Además, logramos que no se tenga que imprimir ningún material, pues el archivo PDF se guarda con los cambios realizados, y se manda por correo para poder ingresar el REA al repositorio. En términos técnicos, la mayoría de programas para abrir PDF permite trabajar con este tipo de formularios, por lo que esta tecnología ha sido fácil de usar por los profesores y no ha dado problemas.
- V. **Subida de los RE al Repositorio Institucional:** Una vez que el profesor ha llenado el formulario PDF, lo envía junto con el archivo del objeto digital al coordinador de su Escuela. El coordinador recopila todo el material de su Escuela, y lo envía periódicamente por correo electrónico a la persona encargada del Repositorio TEC. Cuando el material se recibe, se asigna cada uno de los recursos a un estudiante asistente, el cual ha sido entrenado y está familiarizado con el proceso técnico de envío o subida del material al repositorio. Los estudiantes asistentes permiten que no se forme una cola de espera, al poder repartirse el trabajo y estar creando los REA de forma constante, y también nos permite que los profesores no se tengan que preocupar de hacer un auto depósito en el repositorio.
- VI. **Divulgación y uso por parte de estudiantes:** La etapa final de la estrategia se realiza una vez que el repositorio contenga una cantidad suficiente de REA. En este momento, se debe dar una coordinación entre la VIE y el MEP, para que se comunique a los estudiantes de primaria y secundaria sobre la disponibilidad de los REA, y como pueden acceder a ellos. Mientras más información tengan a disposición los estudiantes de cómo ingresar al repositorio para buscar los recursos, y sobre la importancia de estos, se va incrementar su uso.

Trasversalmente a las etapas de la estrategia de visibilidad se definieron una serie de políticas que soporta y permiten la ejecución de este proceso, las cuales se mencionan a continuación:

- I. **Fortalecimiento del Repositorio Institucional:** Los REA deben quedar almacenados y disponibles de forma pública para que puedan ser utilizados por estudiantes que son externos a la universidad. Claramente, los RE creados son fruto del esfuerzo intelectual de profesores del TEC, por lo que esta producción académica debe ir en el repositorio institucional. Además, la utilización del Repositorio TEC como plataforma tecnológica para los RE es un punto estratégico para ambas partes. El repositorio se va a ver beneficiado con una mayor cantidad de elementos y de visitas. Por otra parte, los estudiantes van a poder contar con una plataforma estable y de fácil uso, que van a poder acceder en cualquier momento y desde cualquier lugar para buscar y utilizar los REA.
- II. **Estándar de metadatos:** El material didáctico creado por los profesores no se puede poner a disposición pública de forma inmediata, pues carece de información adicional o metadatos, para describir todas las propiedades (pedagógicas, descriptivas, etc) de estos objetos digitales. Por esta razón, se decidió utilizar un estándar de me-

tadatos para anotar y enriquecer el material para convertirlo en un REA apropiado. El Learning Object Metadata (LOM) (IEEE Standards Association, 2009) o Metadatos de Objetos de Aprendizaje es un estándar ampliamente usado para describir recursos educativos digitales o objetos de aprendizaje. Aunque LOM es un estándar, no se aplica a contextos específicos por lo que no puede satisfacer necesidades concretas, como en este caso, las necesidades del TEC (Canabal, Sarasa, & Sacristán, 2008). En estas situaciones, existe un mecanismo para adaptar la especificación del estándar que se denomina perfil de aplicación. Según (Duval, Hodgins, Sutton, & Weibel, 2012) un perfil de aplicación es un montaje de diferentes elementos de metadatos seleccionados de uno o más esquemas de metadatos y combinados en uno solo. Su propósito es el de adaptar o cambiar esquemas de metadatos existentes en un paquete específico para las necesidades de una aplicación, mientras mantiene la interoperabilidad deseada con los esquemas bases. Se creó el perfil de aplicación LOM-TEC<sup>1</sup>, el cual define para cada uno de los elementos la repetibilidad, obligatoriedad, momento de generación, y vocabularios adoptados a las necesidades específicas del TEC y de Costa Rica. LOM-TEC además utiliza los vocabularios definidos en (Canabal, Sarasa, & Sacristán, 2008) para que el estándar se pueda aplicar en español. También se extienden algunos vocabularios “desactualizados” para incluir términos más recientes (por ejemplo, el navegador “Firefox”). A partir de este perfil se construyó el formulario editable en PDF para los profesores, y se hicieron las adaptaciones necesarias en el Repositorio TEC para soportarlo. Otro aspecto relevante es que se definieron dos taxonomías básicas para clasificar los REA. La primera, una taxonomía de los niveles educativos de Costa Rica (preescolar, primaria y secundaria). Y la segunda, para clasificar los elementos según la materia o área de conocimiento. A la hora de describir cada REA, LOM permite utilizar taxonomías externas para clasificar el recurso, y se utilizan las dos mencionadas anteriormente para que cada recurso se pueda indexar mejor, y los estudiantes puedan hacer una búsqueda más localizada del material que necesitan.

- III. **Licencia de los REA:** Esta estrategia contempla que los REA puedan tener el mayor alcance posible, y que estén a disposición de estudiantes y docentes para su uso educativo. Para lograr este objetivo, y en concordancia con la definición de REA, se decidió emplear una licencia creative commons. La licencia escogida es “Atribución-NoComercial-CompartirIgual 4.0 Internacional”<sup>2</sup>. Esta licencia permite que cualquier persona pueda hacer uso de los REA, mientras que los autores sean reconocidos, no se use con fines comerciales, y que cualquier adaptación al contenido se comparta de la misma manera.

## 2.2 Aspectos técnicos

Para poder soportar las tareas de la estrategia de visibilidad, fue necesario adaptar el Repositorio TEC para poder subir, buscar y desplegar los Recursos Educativos Abiertos adecuadamente. En nuestro caso, como muchos repositorios latinoamericanos, el repositorio institucional está basado en DSpace<sup>3</sup>. Las modificaciones que se describen a continuación se hicieron sobre DSpace, pero perfectamente se pueden realizar sobre cualquier otro tipo de software, mientras se dedique el tiempo necesario para hacer las modificaciones.

1 <http://hdl.handle.net/2238/6561>

2 <https://creativecommons.org/licenses/by-nc-sa/4.0/>

3 <http://www.dspace.org>



Los cambios más importante que se realizaron fueron:

- I. **Creación de nuevo esquema de metadatos:** Usualmente los repositorios utilizan el esquema de Dublin Core (DC) para describir cada uno de los elementos o envíos, pero este formato no tiene las etiquetas que describen las particularidades de los recursos educativos. Como se mencionó en la sección 3.1, se definió el perfil LOM-TEC para describir los REA, este perfil fue incorporado dentro del repositorio. En el caso de DSpace, la incorporación es sencilla, pues se puede hacer desde la parte administrativa, donde se crea un nuevo perfil y se agregan cada uno de los elementos que son parte él.
- II. **Creación de formulario de envío para los RE:** Cuando se hace un nuevo envío al repositorio, se llena un formulario que corresponde a los metadatos que son necesarios para describir el objeto digital. El Repositorio TEC usa por defecto en todas sus colecciones un formulario para Dublic Core, por lo que fue necesario crear un formulario de envío nuevo adaptado a LOM-TEC. Este formulario del repositorio fue creado en concordancia junto con el formulario editable en PDF, para que los profesores llenaran todos los datos necesarios, y para que se le facilitara el trabajo a los asistentes encargados de crear los REA. En DSpace se modificó el código de la página input-forms.xml, el cual permite definir nuevos formularios de envío.
- III. **Modificación de la vista de envío:** Cuando un usuario visualiza algún elemento en el repositorio, una página muestra sus datos principales: título, descripción, fecha, y autores. En el caso de los REA, hay otros datos que son de relevancia, y que también deben poder visualizar los usuarios para saber si el recurso es de utilidad o no. Estos datos incluyen el tipo de recurso educativo, el destinatario, la dificultad, el nivel educativo para el cual fue creado y el área de conocimiento en la que se clasifica. Adicionalmente, se agregó una opción para visualizar los demás datos de LOM-TEC, para que el usuario pueda acceder de forma fácil a la información técnica y pedagógica del REA. En DSpace estos cambios se registraron en el archivo item-view.xml.
- IV. **Modificación de la búsqueda en el repositorio:** Los repositorios indexan el contenido, y permiten que se realicen búsquedas por ciertos parámetros. Típicamente estos parámetros incluyen el título, el autor, palabras clave y fecha. Como tarea final de la estrategia de visibilidad se tiene el uso por parte de los estudiantes de todos los REA que se están creado, entonces fue necesario modificar estos campos de búsqueda para incluir otros de relevancia. Se modificó la página de búsqueda para agregar los campos de tipo de recurso educativo, el destinatario, la dificultad, el nivel educativo y el área de conocimiento. Esta búsqueda avanzada permite que los usuarios puedan refinar las búsquedas y encontrar los REA que son más apropiados acorde a sus necesidades. El archivo discovery.xml se modificó para agregar nuevos filtros de búsqueda y facetar, correspondiendo a los nuevos campos para los REA.
- V. **Creación del formato de metadatos para OAI:** Finalmente, como un aspecto clave de los REA es que estos sean fáciles de acceder y de compartir, se utilizó el servidor OAI-PMH del repositorio para publicar sus metadatos. Como los REA cuentan con metadatos acordes al perfil LOM-TEC, también fue necesario modificar internamente el código para agregar un nuevo estándar de metadatos, y permitir que se puedan cosechar. Es importante rescatar, que utilizando una tabla de conversión entre LOM-TEC y DC, es posible compartir los REA como si fueran cualquier otro elemento o envío del repositorio, y aunque se pierden los metadatos educacionales, pueden ser cosechados por otros repositorios. Además, se agregaron dos formatos de metadatos, uno para publicar los REA en español, y otro para publicar los metadatos en inglés (para los valores tomados de vocabularios controlados). Técnicamente, en DSpace se modificó el archivo xoai.xml y se agregó el formato LOM-TEC-ES y LOM-TEC-EN. Además se crearon los archivos lom\_tec\_es.xls y lom\_tec\_en.xls.



### 3. Resultados y discusión

El proyecto inició el a finales del primer semestre del 2016, y actualmente se tienen resultados sobre la creación y la aplicación de la estrategia de visibilidad, y se espera que para finales del 2016 e inicio del 2017 se tengan resultados sobre el impacto de dicha estrategia.

De la aplicación de la estrategia podemos mencionar que se han obtenidos varios productos a partir de las tareas planteadas, los cuales se presentan en la Tabla 1.

Productos Productos

Convocatoria inicial Profesores interesados: 11

Escuelas participantes: 7

Material potencial: Mas de 40 apps

educativas, 2 cursos completos en el área de TI, infográficos e interactivos sobre distintos temas, mas de 50 materiales educativos en el área de matemáticas, y otro tipo de material como manuales para hidroponía.

Creación y distribución de un formulario para la creación de RE Formulario editable en PDF con los campos relevantes del perfil de aplicación LOM-TEC.

Estándar de metadatos

Perfil de aplicación completo LOM-TEC, para describir recursos educativos abiertos en el contexto costarricense.

Modificación del Repositorio TEC (plataforma DSpace)

Incorporación del estándar LOM (perfil LOM-TEC), nuevo formulario de envío, modificación de la página de visualización de envíos, modificación de la página de búsqueda, y creación del nuevo formato de metadatos para publicar a través de OAI.

Tarea	Productos
Convocatoria inicial	Profesores interesados: 11 Escuelas participantes: 7 Material potencial: Mas de 40 apps educativas, 2 cursos completos en el área de TI, infográficos e interactivos sobre distintos temas, mas de 50 materiales educativos en el área de matemáticas, y otro tipo de material como manuales para hidroponía.
Creación y distribución de un formulario para la creación de RE	Formulario editable en PDF con los campos relevantes del perfil de aplicación LOM-TEC.
Estándar de metadatos	Perfil de aplicación completo LOM-TEC, para describir recursos educativos abiertos en el contexto costarricense.

Tarea	Productos
Modificación del Repositorio TEC (plataforma DSpace)	Incorporación del estándar LOM (perfil LOM-TEC), nuevo formulario de envío, modificación de la página de visualización de envíos, modificación de la página de búsqueda, y creación del nuevo formato de metadatos para publicar a través de OAI.

**Tabla 1: Tareas y resultados**

Para medir los resultados del impacto de la estrategia de visibilidad, es necesario esperar a que se terminen de crear los REA y se cuente con una cantidad significativa para que se puedan divulgar ampliamente y se usen por parte de los estudiantes. En ese momento, se espera poder medir los resultados con las siguientes métricas:

- I. **Estadísticas de uso de los REA:** El Repositorio TEC permite obtener el número de visualizaciones que tienen cada uno de los elementos disponibles. Tomando en cuenta la cantidad de REA publicados, y la cantidad de visualizaciones de cada uno, se va a poder medir si realmente están siendo utilizados. Además, utilizando Google Analytics<sup>4</sup> se podrá medir si el tráfico del repositorio se incrementó con la incorporación de los REA, y obtener datos demográficos de los usuarios de dichos recursos.
- II. **Encuesta a estudiantes y profesores:** Se planea elaborar una encuesta para que los estudiantes y profesores valoren el impacto y la utilidad de los REA creados. La encuesta se colocará visiblemente en el repositorio, y también se espera aplicarla a estudiantes y profesores de escuelas y colegios donde se haya hecho promoción de los REA (en coordinación con el Ministerio de Educación Pública de Costa Rica).
- III. **Ranking Web de Repositorios:** Una métrica cuantitativa es el Ranking Web de Repositorios<sup>5</sup> que se publica semestralmente y toma en cuenta el tamaño, la visibilidad y el impacto de los repositorios. Si se logra mejorar en el ranking se podría medir el impacto de la incorporación de los REA en el repositorio.

Finalmente, creemos que la aplicación de la estrategia de visibilidad de REA presentada va a tener una serie de beneficios importantes. Se va a contribuir con la educación costarricense al proveer nuevos recursos educativos de calidad que van a poder ser usados tanto por estudiantes como por profesores. El repositorio institucional se va a ver potenciado, al ser el medio de publicación de los recursos educativos, y al incrementar la cantidad y la calidad de los elementos que se encuentran disponibles, vamos a mejorar la visibilidad de la producción académica y científica que se realiza en la universidad. Y como consecuencia del incremento en la cantidad de recursos disponibles en el repositorio, y al ser usado por un nuevo grupo de usuarios, el repositorio se va a consolidar, y posiblemente mejorar en el ranking de repositorios.

#### 4. Conclusiones

Se desarrolló y se puso en marcha una estrategia de visibilidad que permite compartir cualquier tipo de material didáctico de calidad, producto de la docencia e investigación universitaria, en todos los niveles del sistema educativo, a través de Recursos Educativos Abiertos de utilidad para estudiantes de escuelas y colegios. La estrategia contempla el uso del Repositorio Institucional como medio para la publicación de los REA, y permite su uso y reutilización a través de una licencia creative commons. Durante el proceso se

4 <https://analytics.google.com/>

5 <http://repositories.webometrics.info/es>

desarrolló un perfil de aplicación de LOM, conocido como LOMTEC, que permite describir adecuadamente los REA en el contexto educativo para que los estudiantes puedan buscarlos e identificarlos con mayor facilidad. De igual manera, se realizaron cambios importantes en la plataforma tecnológica del repositorio, para que pueda soportar la creación, búsqueda y uso de los REA. Esperamos que la estrategia contribuya con la educación costarricense, de mayor visibilidad a la producción académica y científica realizada en el TEC, y permita que el Repositorio TEC mejore en los rankings. Creemos que la experiencia generada con la creación de la estrategia de visibilidad puede ser replicada por otras instituciones educativas en Latinoamérica que busquen compartir el conocimiento generado a través de Recursos Educativos Abiertos, y que puedan ser de utilidad en sus diferentes realidades y contextos educativos.

## 5. Referencias

- IEEE Standards Association. (2009). 1484.12.1-2002 - IEEE Standard for Learning Object Metadata.
- Canabal, M., Sarasa, A., & Sacristán, J. (2008). LOM-ES: Un perfil de aplicación de LOM. V Simposio Pluridisciplinar sobre Diseño y Evaluación de Contenidos Educativos Reutilizables. Salamanca.
- Vidal Ledo, M., Alfonso Sánchez, I., Zacca González, G., & Martínez Hernández, G. (2013). Recursos educativos abiertos. *Educación Médica Superior*, 27(3), 307- 320.
- UNESCO. (s.f.). Recursos Educativos Abiertos. Recuperado el 1 de 8 de 2016, de Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura: <http://www.unesco.org/new/es/communication-and-information/access-to-knowledge/open%20educational-resources/>
- Butcher, N. (2015). *Guía Básica de Recursos Educativos Abiertos (REA)*. (A. Kanwar, & S. Uvalic-Trumbic, Edits.) Francia: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
- Santos-Hermosa, G., Ferran-Ferrer, N., & Abadal, E. (2012). Recursos educativos abiertos: repositorios y uso. *El profesional de la información*, 21(2), 136-145.
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. (2012). Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4).

# “La producción académica de las Fuerzas Armadas Argentinas visible y abierta al mundo.”

•Autores: *Mónica Silvia Alonso Ramos, Alejandra Carla Castillo*

---

## **RESUMEN:**

En Argentina las Fuerzas Armadas cuentan desde 2012 con un Centro Educativo que agrupa a la Escuela Superior de Guerra; la Escuela de Guerra Naval, la Escuela Superior de Guerra Aérea, el Instituto de Inteligencia de las Fuerzas Armadas y Escuela Superior de Guerra Conjunta. Esta decisión ha hecho que además de compartir el mismo ámbito, biblioteca y actividades conjuntas se fomente la interoperabilidad entre las distintas Fuerzas. Ahora ha llegado el momento de hacer visible su producción al mundo y a partir de diciembre de 2013, con la promulgación de la Ley 26.899, sobre Repositorios Digitales, el CEFADIGITAL abre una ventana al mundo para dejar ver la amplia producción casi desconocida de las Fuerzas Armadas, elaborada en estas instituciones desde principios del siglo XX.

La implementación del Repositorio Digital garantiza el acceso abierto a la información generada por todos los integrantes de las instituciones educativas, alumnos, docentes, investigadores y de los diferentes ámbitos y especializaciones. La visibilidad de la producción académica y científica permite que se comparta el conocimiento y contribuya a su crecimiento continuo tanto en el ámbito nacional como internacional.

## **ABSTRACT:**

Since 2012, the Argentine Armed Forces have an Educational Centre in which the Army Staff College, the Navy Staff College and the Air Force Staff College, Intelligence College and the Joint Staff College carry out their activities. This decision has allowed colleges not only to share facilities, library and activities, but also to promote interoperability among the different Forces. It is now time to make their production visible worldwide. Therefore, since 2013 with the enactment of Law 26899 of Digital Repositories, CEFADIGITAL gives the chance to show their academic results to the world.

The implementation of the Digital Repository guarantees Open Access to the information produced by all members of these Institutions: students, professors, and researchers of different fields of specialization. Making academic and scientific production visible provides with the possibility to transfer knowledge and to contribute to its permanent growth both at national and international levels.

## INTRODUCCIÓN:

El planteo del Acceso Abierto tiene como finalidad la democratización de los resultados de las investigaciones. De acuerdo a la historia de los últimos cincuenta años en América Latina, parecería que palabras como DEMOCRACIA – ACCESO ABIERTO, no estuvieran relacionadas con la idea que se tiene de las Fuerzas Armadas.

Cuando se habla de interoperabilidad como la capacidad que poseen varios sistemas heterogéneos de intercambiar información e interpretarla de la misma manera es fácil relacionar el término con la forma de actuar de las Fuerzas Armadas ya que es de suma importancia este concepto en especial en el marco de las organizaciones multinacionales. Para poder operar de forma eficiente se requiere de la interoperabilidad, lo que implica la solución de grandes problemas de entendimiento y coordinación en todos los niveles, y comprende tanto estructuras de gobierno y militares como organizaciones civiles de apoyo.

Se puede hablar de alianzas o coaliciones dentro de las operaciones multinacionales dependiendo del lugar de la acción pero hay que tener en cuenta que cada país tiene sus propios intereses, cultura, idioma, capacidades, equipamiento que muchas veces no son compartidos con el resto.

Pero si bien las necesidades y condiciones son muy diferentes, se trata el mismo problema y eso hace que haya una necesidad de cooperación y alianza entre países que lleva a compartir información. En este modelo de interoperabilidad conceptual lo importante es que todos los sistemas intervinientes tengan la misma percepción de lo que está sucediendo, que tanto las tropas como los técnicos informáticos entiendan lo mismo, ya sea en el terreno como frente a una pantalla manejando una aplicación, los datos y la percepción de los mismos no varíe, lo importante es que la forma de entender sea la misma.

Estos conceptos que se aplican en la vida diaria y dentro del ámbito de las Fuerzas Armadas de los diferentes países se utilizan ahora en el marco de los repositorios digitales. La raíz es la misma y a pesar de los diferentes idiomas, cultura, intereses, la interoperabilidad permite la disponibilidad e interpretación de los datos con un mismo entendimiento.

En Argentina, una de las medidas que se tomó en 2012 por Resolución MD N° 1427/06, para promover la interoperabilidad interna, fue la creación del Centro Educativo de las Fuerzas Armadas, en el cual funciona la Escuela Superior de Guerra; la Escuela de Guerra Naval, la Escuela Superior de Guerra Aérea, la Escuela Superior de Guerra Conjunta y si bien no comparte el predio el Instituto de Inteligencia de las Fuerzas Armadas.

Pero una de las asignaturas pendientes de nuestras Fuerzas Armadas era la visibilidad. Dentro de sus instituciones educativas se ha creado, a lo largo de los años, una gran cantidad de producción científica, académica, tecnológica que durante muchos períodos fue accesible solo al centro de origen, lo que conlleva una limitación importante para la continuidad de las investigaciones, en algunos casos, muy especializada.

Cabe recordar que la Escuela Superior de Guerra del Ejército fue fundada en el año 1900, la Escuela de Guerra Naval ha cumplido ya ochenta y dos años y la de Fuerza Aérea fue

creada en 1944. En sus aulas se capacitaron historiadores, comandantes, investigadores, tanto militares como civiles, argentinos y extranjeros.

Los nuevos requerimientos de información de carreras con estudios especializados como ciencias políticas y relaciones internacionales resulta en que la producción académica de estas escuelas sea de mayor interés, pero la falta de difusión hace que la comunidad vea estos recintos como velados o simplemente desconozca la calidad y cantidad que se produce a diario en nuestras instituciones.

En el mundo, los repositorios digitales van en aumento, basta visitar la página del Registry of Open Access Repositories y del Directory of Open Access Repositories para ver cómo año tras año se incrementan los repositorios, los objetos digitales y las visitas a los mismos.

Los centros educativos de los diferentes países están primeros en los rankings, pero no sucede lo mismo con los repositorios de las Fuerzas Armadas.

España cuenta con el Centro Superior de Estudios de la Defensa Nacional que cuenta con monografías que se publican desde el año 1990, conferencias, actividades, revistas y boletines que se pueden descargar en formato pdf. Pero su configuración no permite las búsquedas y recuperación de información con las que cuenta un repositorio digital.

En América del Sur, Ecuador cuenta con un repositorio digital realizado con el software DSpace que agrupa varias comunidades, distintas escuelas que pertenecen a la Universidad de las Fuerzas Armadas. (<http://repositorio.espe.edu.ec/handle/21000/3251>) y cuenta con archivos de video y de imágenes además de los de texto.

En Argentina, a partir de diciembre de 2013, con la promulgación de la Ley 26899, sobre Repositorios Digitales, surge la obligación, no solo la necesidad, de que aquellas instituciones públicas o investigadores, tecnólogos, docentes, becarios de posdoctorado y estudiantes de maestría y doctorado por cuya actividad de investigación reciban financiamiento del Estado nacional, “deberán desarrollar repositorios digitales institucionales de acceso abierto, propios o compartidos, en los que se depositarán la producción científico-tecnológica resultante del trabajo, ... de sus investigadores, tecnólogos, docentes, becarios de posdoctorado y estudiantes de maestría y doctorado” (Argentina, 2013, art. 1).

Esta imposición ha actuado como disparador para que el Centro Educativo de las Fuerzas Armadas ponga en marcha el proyecto del CEFADIGITAL.

La Biblioteca Grl. Belgrano, biblioteca del Centro Educativo de las Fuerzas Armadas se encuentra integrada física y funcionalmente por las tres bibliotecas de las escuelas de guerra específicas: Ejército, Naval, Aérea, así como la de Guerra Conjunta. La biblioteca del Instituto de Inteligencia de las Fuerzas Armadas se encuentra en las instalaciones que ocupa dicho instituto y su integración funcional, se logró a través del repositorio digital.



A partir de la creación de un Repositorio Institucional propio, de acceso abierto e interoperable, el Centro Educativo de las Fuerzas Armadas se ha convertido en referente de otras instituciones educativas militares.

En las cinco escuelas del Centro Educativo de las Fuerzas Armadas, los estudiantes realizan sus trabajos finales de carrera y gran parte de los docentes elaboran distintos trabajos académicos y científicos que no siempre tienen visibilidad, ni posibilidad de publicarse en otros ámbitos por fuera de la institución. Actualmente, toda esta producción carece de un registro único, se encuentra dispersa y su acceso es parcial y, en algunos casos, no es posible tal acceso ya que no existe una base de datos ordenada ni completa de los diversos materiales que se producen como tampoco de sus archivos digitales.

Los trabajos finales realizados por los alumnos son un material de consulta muy importante para todos aquellos que quieran desarrollar nuevos trabajos sobre temas afines o bien para los que necesiten o se interesen por las temáticas que se han abordado, sean o no integrantes de la comunidad educativa de las FFAA.

Los libros escritos por los docentes en ejercicio de sus cátedras son denominados institucionalmente Contribuciones Académicas e integran la bibliografía de referencia de los programas de las asignaturas y por ello es necesario que los estudiantes y otros docentes o interesados puedan consultar y tener acceso a ellos. Lo mismo sucede con los artículos de las revistas institucionales que se producen en el marco de los institutos que integran el CEFADIGITAL. En este sentido, la creación de un repositorio institucional es la mejor alternativa y el recurso tecnológico más apto para acceder a esos materiales.

El objetivo de este trabajo es exponer el proceso de gestación del repositorio institucional del Centro Educativo de las Fuerzas Armadas, su implementación actual y las vías de desarrollo planteadas.

## **MATERIALES Y METODOLOGÍA:**

La técnica empleada es la del análisis documental que permitió disponer de criterios teóricos para poder formular y diseñar el proyecto. El método de análisis-síntesis se utilizó para formar la estructura de las bases metodológicas de implementación del repositorio, se revisó además la normativa para el acceso abierto, y directrices para su creación, se consultó la bibliografía producida por la Universidad Nacional de La Plata, como también el código de buenas prácticas propuestas por el Consejo Superior de Investigaciones Científicas, del gobierno de España. También se tomó contacto con la Biblioteca Nacional de Maestros del Ministerio de Educación de la Nación que cuentan con un repositorio institucional que almacena información de todas las áreas que lo conforman. Además, se buscó el asesoramiento de la Universidad Nacional de La Plata, quienes han compartido sus prácticas y experiencias desinteresadamente y de manera irrestricta, a través del Servicio de Difusión de la Creación Intelectual (SEDICI).

Como herramienta de administración y gestión de los contenidos que conforman el Repositorio se tomó la decisión de elegir el Dspace versión 4.2., que cumple con la normativa vigente por ser una plataforma de software libre, también recomendado por el Servicio de Difusión de la Creación Intelectual de fácil configuración de funcionalidades

que respondan a las necesidades específicas de cada organización por lo que cada área puede definir el flujo de trabajo y proceso de publicación de sus documentos de forma independiente.

Esta plataforma es utilizada por más de mil cuatrocientos repositorios en el mundo, lo que le da fiabilidad y se puede considerar amigable para la introducción de metadatos.

Su estructura permite organizar la información en comunidades que a su vez, se segmentan en colecciones de documentos de acuerdo a sus necesidades. Este era un requisito indispensable para la Institución dado que se trabaja conjuntamente en las distintas Escuelas pero respetando las clasificaciones de las colecciones en cada una de ellas, cada Escuela toma decisiones en forma independiente dentro de su comunidad y cada colección de documentos puede contar con restricciones de acceso personalizadas.

Además, otra ventaja es que almacena casi cualquier tipo de formato y documento, utilizando para la catalogación de los mismos el estándar Dublin Core. En las instituciones educativas se presentan diferentes tipos de formatos, si bien el más común es el de texto, se incorporarán en breve imágenes y videos de conferencias, ya que varias veces al año cada Escuela tiene un ciclo de conferencias, ejemplo: Seminario de Malvinas en la Escuela de Guerra Naval y o ciclos de conferencias que responden a distintas temáticas.

El sistema de búsqueda permite utilizar filtros para delimitar búsquedas por comunidad, colección, autor, título, año, temática, área de interés, clasificación, entre otras, e incluso localizar información dentro del texto completo de los documentos.

Se organizaron los flujos de trabajo y la responsabilidad de los participantes. Para la realización del proyecto se convocaron a civiles y militares, en actividad y en situación de retiro, con diferentes formaciones académicas y experiencias profesionales en distintas áreas como Derecho, Traductología, Informática, Educación y Bibliotecología entre otros. El equipo de trabajo multidisciplinario perteneciente a la Escuela Superior de Guerra Conjunta fue coordinado por el subdirector de la Escuela.

Definido y conceptualizado de esta manera el grupo, se implementó un método, que permitió prácticas definidas, útiles y eficientes al considerar los esfuerzos de sus integrantes obteniendo un mayor aprovechamiento de los recursos institucionales.

Posteriormente, una vez concebido el proyecto del Repositorio Institucional se capacitó a los distintos responsables de las Escuelas intervinientes. Teniendo en cuenta que "... las bibliotecas universitarias son el estamento que por su experiencia en la gestión de la información en todas sus formas y el contacto con el conocimiento, deberá liderar la implementación de los repositorios institucionales con el fin de lograr la competitividad educativa..." (Molina Piñeiro et al, 2015), se determinó que la administración de cada comunidad esté a cargo de equipos en los que participen los bibliotecarios, quienes además mantendrán comunicación con las diferentes dependencias institucionales para la gestión de los contenidos a subir, contacto con los autores, así como promover su uso y los servicios que se derivan de él.

De todas las instituciones educativas de las Fuerzas Armadas el CEFADIGITAL está por el momento integrado por los siguientes institutos Escuela Superior de Guerra, Escuela

de Guerra Naval, Escuela Superior de Guerra Aérea, Escuela Superior Técnica, Instituto de Inteligencia de las Fuerzas Armadas y Escuela Superior de Guerra Conjunta, con una cantidad aproximada de cincuenta tesis promedio elaboradas por año por institución.

Además, cuentan con carreras de posgrado abiertas a toda la sociedad argentina y de extranjera. Cada una de ellas tiene particularidades específicas, y se encuentran en diferentes etapas del proceso.

Dentro de cada Escuela se ha convocado a una comisión de trabajo para que realice una adecuada curación de contenidos evaluando la calidad y viabilidad de las publicaciones. La Comisión está formada por personal militar y civil, expertos en las diferentes áreas de producción científica y académica.

Se ha puesto especial atención a la estética del repositorio, según Bueno et. al. (2011) “La evaluación podrá basarse en el empleo de medidas cualitativas y cuantitativas. La calidad de los repositorios, y por tanto, de sus recursos, deberá medirse por el rigor, la imparcialidad, la actualidad, cobertura y complejidad de los conocimientos que intenta comunicar, por su vocabulario y por la calidad de los recursos empleados (como textos, imágenes, videos, etc.) Pero también por su calidad didáctica, su eficacia, su capacidad de motivación, su diseño instruccional, su adecuación a los estudiantes, su enfoque creativo y aplicativo de las actividades, etc. Y por último, por su calidad técnica y estética entre lo que debe destacar su facilidad de uso o su forma de gestionar las interacciones entre el material y el usuario”.

El desarrollo del logo como concepto señala el paso del papel a lo digital, del enfoque tradicional de la difusión del conocimiento a las nuevas alternativas que nos presenta la web 2.0.



## **RESULTADOS PARCIALES O FINALES:**

El repositorio institucional CEFADIGITAL ha sido puesto en marcha oficialmente en octubre de 2015 y hasta el momento en el repositorio se han subido alrededor de doscientos cincuenta objetos digitales. Para octubre se esperan tener, todas las colecciones definidas y quinientos documentos subidos a la misma.

En un año las comunidades contarán con un total de entre setecientos y mil documentos subidos al repositorio, constituyendo así el mayor repositorio de Fuerzas Armadas en Latinoamérica.

Las comunidades tienen distinta problemática a la hora de subir materiales al repositorio,

mientras que la Escuela Superior de Guerra Conjunta, por ser una institución de reciente creación, trata objetos que son nativos digitales, la Escuela de Guerra Naval, con más de ochenta años de tesis, contribuciones académicas, revista y libros, deberá iniciar un proceso de digitalización.

Dado que muchas de las contribuciones académicas, tesis, revistas sólo se producían en papel, esto obliga a realizar diferentes procesos de digitalización, no sólo con el objetivo de subirlo al repositorio sino como preservación de los documentos. Esto requiere de un proceso de análisis global y luego particular, es decir, una vez determinadas las colecciones, se requiere el estudio de los datos, cantidad de documentos, intervenciones requeridas, prioridades, procesos y recursos para ponerlos en práctica. Se pretende subir los documentos digitalizados e instrumentar herramientas como el reconocimiento óptico de caracteres para permitir la búsqueda rápida en los textos.

Como casos particulares podemos mencionar los siguientes: uno de los desafíos que se plantea la Escuela Superior de Guerra (Ejército) es la digitalización de su “colección histórica” que data del siglo XXVIII, lo cual constituye la mayor riqueza de su acervo bibliográfico. Algo similar ocurre en la Escuela Superior Técnica con la colección de la revista institucional, sobre tecnología e ingeniería militar, que se editó del año 1951 al 2008, en formato papel, que es de excelente nivel académico. En otro orden uno de los objetivos de la Escuela Superior de Guerra Conjunta es hacer visible a los autores de Contribuciones Académicas y artículos de la revista institucional Visión Conjunta publicando el material en idioma español y llegar también a otro público en idioma inglés, ampliando así su visibilidad en el mundo, lo cual se logra a través del repositorio institucional.

## **CONCLUSIONES:**

La aparición de la Web 2.0, como medio principal de intercambio de información, ha hecho que cualquier individuo pueda tener acceso a la producción científica proveniente de instituciones universitarias, centros de investigación y academias en todo el mundo. Esta producción es uno de los indicadores importantes en el momento de evaluar el grado de excelencia y desarrollo de la generación de conocimiento de una institución o país determinado (Machín, Lugo & Díaz, 2013)

La gestión de materiales digitales de enseñanza y aprendizaje constituye uno de los retos más importantes a los que desafían las bibliotecas universitarias, pues su éxito obedecerá esencialmente del nivel de adopción y uso que haga de ellos la comunidad académica, especialmente los docentes. Las bibliotecas, deben asumir la función de crear las condiciones adecuadas que permitan orientar, tratar, almacenar y dar visibilidad a los materiales producidos para la enseñanza y la investigación.

Por otra parte, los distintos integrantes de los institutos que componen el CEFADIGITAL han comenzado a tomar conciencia de la importancia de visibilizar sus resultados de investigación mediante la publicación de sus artículos en el repositorio institucional, facilitando la recolección de información a nivel institucional.

El mejoramiento de la visibilidad de la investigación es una labor permanente, que se irá concretando con la incorporación de nuevos profesionales al grupo de trabajo original

con los aportes de las nuevas instituciones adheridas al CEFADIGITAL.

De esta manera, se continuaría centralizando la producción intelectual del CEFFAA en el repositorio institucional, el cual concentraría las tres grandes áreas que la producen: docencia (material didáctico para estudiantes y profesores); investigación (resultados de investigación); y editorial (revistas, libros digitalizados y electrónicos).

Esta decisión institucional constituye una de las estrategias más viables para garantizar el acceso abierto a la información que se produce en su claustro académico contribuyendo a incrementar la visibilidad de la producción académica y científica de docentes, investigadores, alumnos y egresados, permitiendo a la vez cumplir la misión de compartir los conocimientos con la sociedad y con otras comunidades educativas del ámbito nacional, regional e internacional.

*“La biblioteca es destinada a la ilustración universal y más poderosa que nuestros ejércitos para sostener la independencia”*

Gral. José de San Martín

## **Bibliografía**

Benítez de Vendrell, B. (2014). Manual de metadatos para el repositorio Argos (1° ed., pp. 1-63). Posadas, Universidad Nacional de Misiones. Facultad de Humanidades y Ciencias Sociales. Secretaría de Investigación y Posgrado.

Bueno de la Fuente, G. (2011). Estrategias para el éxito de los repositorios institucionales de contenido educativo en las bibliotecas digitales universitarias. Bid. Consultado el 3 de julio de 2016, de <http://bid.ub.edu/26/bueno2.htm>

Bustos González, A.; Fernández Porcel, A. (2008). Directrices para la creación de repositorios institucionales en universidades y organizaciones de educación superior internacional. Bruselas, Comisión Europea EuropeAid, Oficina de Cooperación Consultado el 7 de julio de <http://repository.urosario.edu.co/handle/10336/223>

Centro Superior de Estudios de la Defensa Nacional (CESEDEN). Consultado el 07 de julio 2016, de <http://www.defensa.gob.es/ceseden/>

Creative Commons. Obtenido el 4 de julio de 2016 de <http://creativecommons.org/>

Dublin Core Metadata Initiative. Consultado en junio de 2016 de <http://www.dublincore.org/>

Escuela Superior de Guerra Conjunta (2014). Proyecto: Creación de un repositorio digital institucional de acceso abierto e interoperable “CEFADIGITAL” en el centro educativo de las fuerzas armadas. Buenos Aires: ESGC.

Machín, M. P., Lugo, Y. G., & Díaz, M. del P. F. (2013). Análisis bibliométrico sobre la visibilidad científica y el factor de impacto en ciencias de la educación. EDUMECENTRO, 3(3), 99-112.

Molina Piñeiro, M., Marrero Sera, E. & Puentes Puente, A. (2015). Los repositorios de acceso abierto como alternativa para la visibilidad de la ciencia en las universidades: estudio de caso. Revista Cubana de Información en Ciencias de la Salud, 26(4) Recuperado en 04 de agosto de 2016, de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2307-21132015000400003&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132015000400003&lng=es&tlng=es).

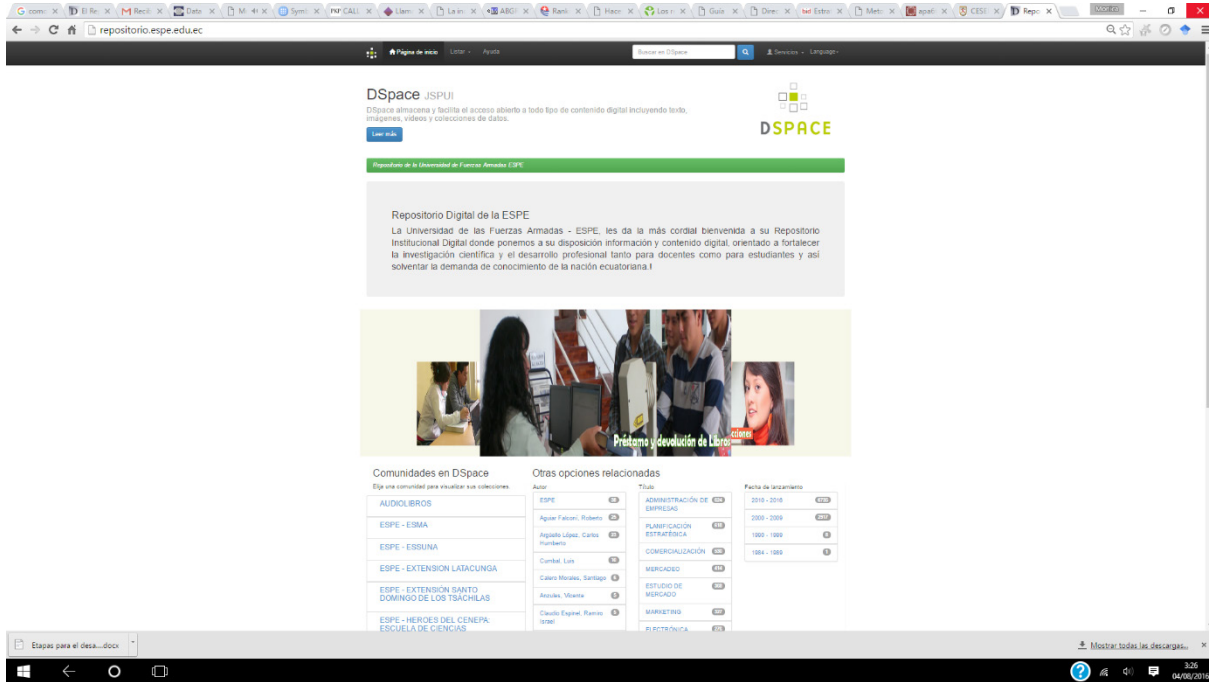
Red Federada de Repositorios Institucionales de Publicaciones Científicas Recuperado el 6 de junio de 2016 de <http://www.lareferencia.info/vufind/>

Sistema Nacional de Repositorios Digitales (SNRD). Recuperado el 7 de julio de 2016 de <http://repositorios.mincyt.gob.ar/>

The Directory of Open Access Repositories – OpenDOAR. Recuperado el 6 de Julio de 2016 de <http://opendoar.org/index.html>



# ANEXO 1: Portal Repositorio de la Universidad de las Fuerzas Armadas de Ecuador



# ANEXO 2: CEFADIGITAL





# Panel

## Implementación de ORCID en repositorios institucionales

Caterina G. Pavão (UFRGS)

# “Permitiendo la interoperabilidad de la información con identificadores”

*Lilian Pessoa, Director Regional ORCID*

Se describen las características y funcionalidad de la iniciativa de Acceso Abierto ORCID (Open Researcher and Contributor ID) para proveer un identificador digital persistente a los autores o investigadores que les permita crear conexiones entre ellos, sus contribuciones y afiliaciones. Se describe el proceso de registro y creación del perfil ORCID ID a nivel investigador así como, las ventajas para las organizaciones que adquieren la membresía institucional, entre las destacan:

- 1) Reducir/eliminar la necesidad de desambiguación de los investigadores dentro y entre sistemas;
- 2) Obtener información completa sobre las actividades de sus investigadores en sus sistemas;
- 3) Permitir actualizaciones automáticas sistema-sistema para reportes de investigadores;
- 4) Obtener información sobre investigadores visitantes y colaboradores;
- 5) Hacer interoperable el repositorio institucional en los registros de los investigadores;
- 6) Mayor visibilidad, incluso internacional, a la producción de la institución;
- 7) Validar la afiliación de su investigador, reconocimientos, o trabajos, construyendo la confianza en la infraestructura general de la investigación;
- 8) Mejorar la velocidad y precisión para reportar la investigación;
- 9) Mantener conexiones con investigadores sin importar cambios en su nombre/afiliación;
- 10) Asegurar que sus investigadores cumplan con sus fundadores y editoriales que soliciten registro ORCID;
- 11) Aprender de las experiencias de otras organizaciones y compartir las suyas;
- 12) Ser reconocidas públicamente por sus buenas integraciones con ORCID;
- 13) Involucrarse y ayudar a ORCID a trazar sus metas y prioridades futuras al formar parte de la comunidad ORCID.

# “Estandarización de identificadores de investigadores con ORCID de la Universidad del Rosario de Colombia para la visibilidad y posicionamiento en web”

*Malgorzata Lisowska, Universidad del Rosario*

Se describe la experiencia del Centro de Recursos para el Aprendizaje y la Investigación (CRAI) de la Universidad del Rosario de la ciudad de Bogotá, Colombia, en la implementación de ORCID para la identificación de autores ante editoriales y organismos de financiación, así como para la unificación de la producción académica y fortalecimiento de la visibilidad de la producción de los investigadores de su Universidad. El proyecto de creación de identidad digital inicia en diciembre de 2015 con la adquisición de la membresía institucional ORCID y tres meses después, se lleva a cabo la normalización de 8 perfiles e identidad web desarrollado por dos profesionales durante 6 meses de un total de 408 investigadores. Posteriormente, se lleva a cabo la integración de ORCID con los sistemas existentes DSpace – OJS – VIVO – PlumX - PURE.

Los resultados obtenidos fueron la creación de 386 (94.61%) perfiles de investigadores normalizados, el 85.7% de investigadores ranqueados gracias a la normalización de la identidad digital y un mayor impacto en la visibilidad de los investigadores y su producción. Como estrategias futuras se plantean la actualización de 5 perfiles principales (ORCID, PlumX, Google Scholar, Scopus ID y Research ID) y perfiles opcionales de otros índices de revistas y plataformas.

# “Resultados de ORCID en la UAEMex dos años después”

*Veronica Pichardo, Universidad Autónoma del Estado de México*

Se describe la experiencia de la Red de Revistas Científicas de America Latina y El Caribe, España y Portugal (REDALYC) de la Universidad Autónoma del Estado de México, en la implementación de ORCID con su plataforma de Autores REDALYC. Esta plataforma permite a un autor identificar sus trabajos publicados en revistas indizadas en Redalyc, crear su página de autor, visualizar los indicadores de publicación, co autoría y descargas, así como también, podrá crear una cuenta de ORCID e integrar su producción en otras bases de datos.

# Talleres

# Taller 1: “El ciclo de vida de los datos de investigación desde la perspectiva de la ciencia en abierto”

*Instructor: Dra. Remedios Melero, CSIC, España.*

Doctora en Ciencias Químicas, por la Universidad de Valencia. Trabaja en el Instituto de Agroquímica y Tecnología de los Alimentos del Consejo Superior de Investigaciones Científicas de España (CSIC) como científico titular y es editora de la revista *Food Science and Technology International*. Vicepresidenta de *European Association of Science Editors* hasta el año 2012. Es profesora en el master de Documentación Digital impartido por la Universidad Pompeu Fabra y miembro del comité científico de Redalyc y de Scielo España. Sus temas de trabajo incluyen publicaciones electrónicas, procesos de arbitraje académico, acceso abierto a la producción científica y propiedad intelectual. Es miembro del grupo de trabajo Acceso Abierto a la Ciencia para el estudio de revistas, repositorios y políticas institucionales de acceso abierto. Ha participado en el proyecto europeo NECOBELAC para la difusión y creación de redes en torno a la publicación científica y el acceso abierto a la ciencia, entre Latinoamérica y Europa. Actualmente trabaja en un proyecto nacional coordinado con la Universitat de Barcelona sobre el grado de implantación y sostenibilidad de recursos en acceso abierto, y hasta julio de 2016 ha participado en el proyecto europeo FOSTER (Facilitate Open Science Training for European Research).

## **Objetivos:**

Entender la relevancia de preservar y compartir los datos de investigación con el resto de la comunidad científica.

Aprender a crear un plan de gestión de datos

## **Temas:**

Significado de la Ciencia en Abierto (Open Science)

Los datos de investigación tipología y formatos

Características de los datos de investigación en abierto

Etapas del ciclo de vida de los datos de investigación: Colección, procesado, análisis, preservación, localización y reutilización

Plan de gestión de datos: cómo crear un plan de gestión de datos

Localización de datasets

Citación de datos

Licencias para datos

Repositorios de datos

Revistas de datos

Políticas de acceso abierto



# Taller 2 “La gestión de la propiedad intelectual en el entorno digital”

*Instructor: Dra. Mag. Margarita Lisowska CRAI-Biblioteca UR, Universidad del Rosario Colombia.*

Magister en Bibliotecología e información científica, de la Universidad Jagiellona de Cracovia - Polonia Profesional con mas de 20 años de experiencia en gestión y procesos de cambio en las bibliotecas universitarias.

Desde 1999 - Directora del CRAI de la Universidad del Rosario en Bogotá, Colombia  
2010-2013 - Coordinadora técnica del proyecto nanciado por BID “LA Referencia”  
2007-2010 - Miembro fundador de la Biblioteca Digital Colombiana

## **Contenido:**

Conceptos básicos de Derecho de Autor  
Diferencias entre Derecho de Autor y Copyright  
Acceso abierto versus derecho de autor  
Producción intelectual en las universidades y su difusión en la web  
Autoarchivo en los repositorios institucionales

# Taller 3 “Conguración de directrices OpenAire 3.0 y uso de vocabularios controlados COAR en el marco europeo y latinoamericano”

*Instructor : Mtro. Antonio Felipe Razo, REMERI, México.*

Profesor de la carrera de Diseño de Interacción y Animación de la Universidad Iberoamericana en Puebla, actualmente es el Coordinador Técnico de la Red Mexicana de Repositorios Institucionales REMERI(<http://www.remeri.org.mx>). Colaborador y consultor del proyecto REMERI de CONACyT desde el 2012 como responsable del desarrollo y mantenimiento de la plataforma técnica de REMERI denominada INDIXE.

Consultor para la primera etapa del proyecto LA-Referencia (<http://www.lareferencia.info>) del Banco Interamericano de Desarrollo y RedClara en el 2013. Ingeniero en Sistemas Computacionales, Maestro en Ciencias y Maestro en Diseño de Información por la Universidad de las Américas Puebla UDLAP. Participante del “Programa de capacitación sobre Repositorios Institucionales” de RedCLARA en el 2013 (4 módulos) y de los “Cursos de Administración de DSpace” para CUDI por Arvo Consultores (España) en 2014. Impartió el taller “Fundamentos Técnicos para un Repositorio Institucional en el Marco del Repositorio Nacional” en la reunión de primavera de CUDI en mayo del 2016 y el curso de “Tecnologías para el Acceso Abierto” para la Universidad de Guadalajara en enero de 2016.

## **Contenido**

Directrices, vocabularios y consideraciones

Configuración de formularios en DSpace y servidor de metadatos

Validación de servidores de metadatos y registro como proveedor OpenAIRE 3.0

# Taller 4 “Cómo construir una identidad académica digital”

*Dr. Enrique Orduña-Malea, UPV, ESPAÑA*

Es Ingeniero Técnico de Telecomunicaciones, Licenciado en Documentación, Master en contenidos multicanal y Doctor en Documentación (Tesis doctoral orientada a la cibermetría académica) por la Universidad Politécnica de Valencia (UPV).

Es autor de más de 50 artículos y capítulos de libro académicos y profesionales, entre las que se incluyen más de 25 publicaciones recogidas en los Journal Citation Reports (JCR). Es co-autor de las monografías “Cibermetría: midiendo el espacio red” y “La revolución Google Scholar”, de inminente publicación. Desde 2008 es miembro del Grupo ThinkEPI (Grupo de Análisis sobre Estrategia y Prospectiva de la Información); desde 2012 del Grupo de investigación EC3 (Evaluación de la Ciencia y de la Comunicación Científica) de la Universidad de Granada (UGR); y, desde 2014 es miembro fundador del Grupo Trademetrics, orientado a la aplicación de indicadores web en el análisis de la presencia e impacto de las empresas y marcas en la Web.

## **Objetivos:**

COMPRENDER los cambios en los modelos de comunicación y evaluación académica.  
CONOCER distintas acciones y estrategias para mejorar la visibilidad académica online de una universidad a través de la visibilidad de sus investigadores.  
CREAR y OPTIMIZAR perfiles académicos en la Web.

## **Temas:**

Metamorfosis en la comunicación y evaluación científica.  
Estrategia general de comunicación:  
Creación de perfiles académicos creados desde:  
Bases de datos bibliográficas: ORCID, ResearcherID.  
Repositorios, tanto institucionales como temáticos (ej. REPEC).  
Motores de búsqueda académicos: Google Scholar Citations.  
Redes sociales académicas: Academia.edu, ResearchGate, Mendeley.  
Redes sociales profesionales: LinkedIn.  
Servicios contenedores: ImpactStory, Scholar Mirrors

# Posters



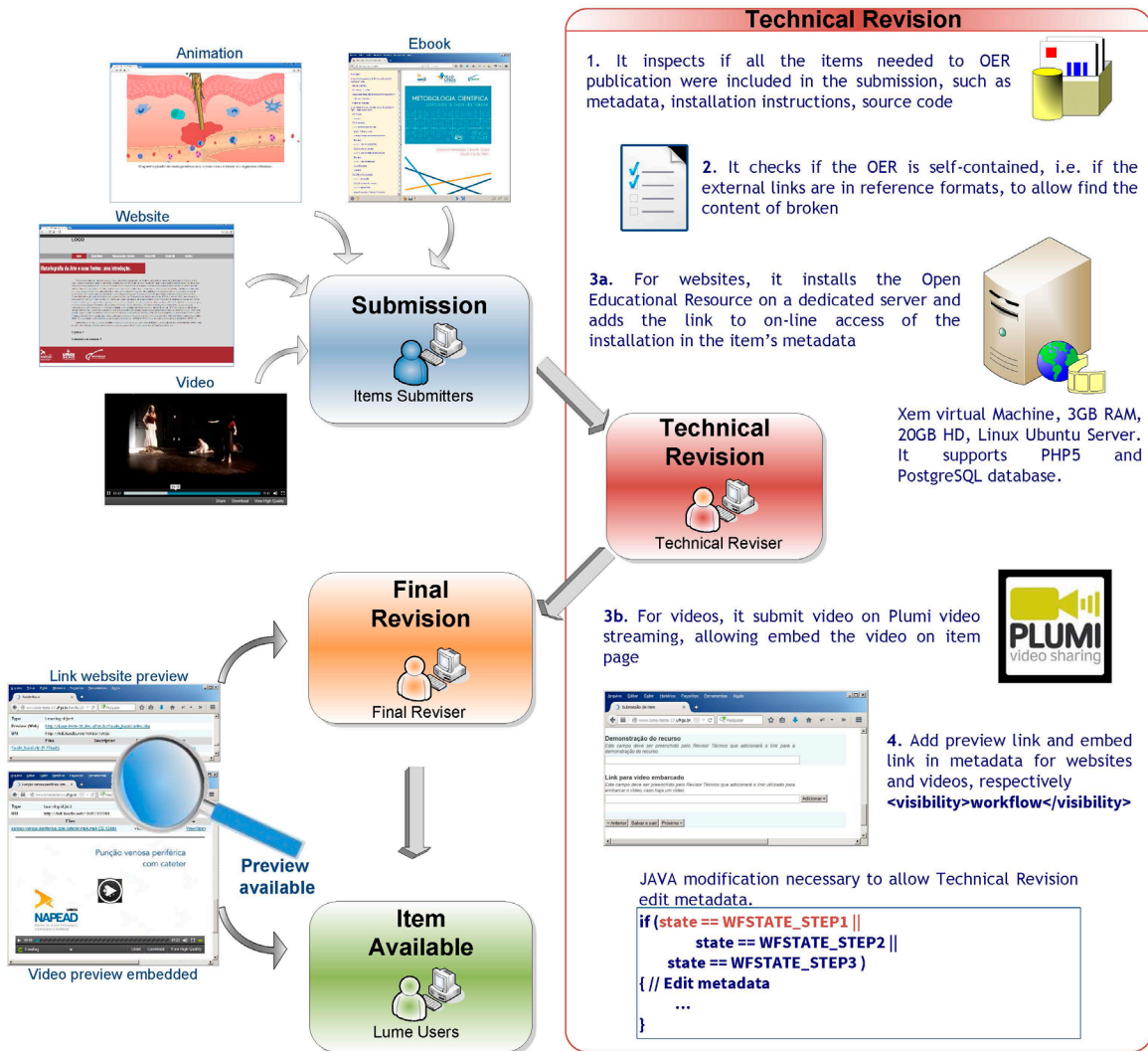
# An Implementation of Technical Revision in DSpace Allowing Open Educational Resource Browser Access



Manuela Klanovicz Ferreira, Zaida Horowitz,  
Janise Silva Borges da Costa, Caterina Groppo Pavão  
Centro de Processamento de Dados - Universidade Federal do Rio Grande do Sul (UFRGS)  
Porto Alegre, RS, Brazil E-mail: [lume@ufrgs.br](mailto:lume@ufrgs.br)

## Abstract

This work shows how the DSpace software was changed to create the technical revision step in the workflow submission of the Open Educational Resources (OER) community in Lume - Digital Repository of Federal University of Rio Grande do Sul. The main goal of technical revision step is allow access the OER directly on browser. OER installation is not required.



## Conclusion

On Technical Revision, a verification of OER's technical properties and the inclusion of link to its installation on-line access allow the others reviewers access it without the need to install it, using only the browser, streamline the submission process. This implementation also offers an OER's preview to the end users, facilitating the access and evaluation to its utilization. In this way, we intend to make Open Educational Resources as accessible as possible.



International Conference BIREDIAL-ISTEC 2016 - Mexico, October 17 - 19



[www.lume.ufrgs.br](http://www.lume.ufrgs.br)

Rua Ramiro Barcelos, 2574  
90035-003, Porto Alegre, RS, Brazil  
Phone: +55 513308-5071

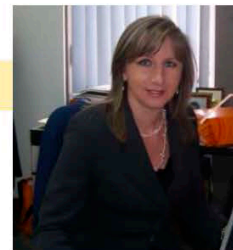






## PERFIL ACADÉMICO WEB

### Malgorzata Lisowska-Navarro



Ha culminado con éxito la normalización de su presencia web.

 **Para acceder a su perfil, haga click en cada ícono.**

#### ORCID

Use su ORCID ID para identificarse como autor ante editoriales y organismos de financiación.



Conozca sus índices de autor y citas, su ranking en Webometrics y reciba notificaciones sobre nuevas citas a su producción científica.

#### RESEARCHERID

Conozca qué autores citan sus documentos indexados en ISI Web of Science.

#### ResearchGate

Amplíe sus redes académicas, comparta sus publicaciones y participe en espacios de discusión.



Conozca las métricas de sus artículos publicados en revistas indexadas en Redalyc.

#### irelis

Conozca y/o modifique su firma como autor científico.



Infórmese acerca de oportunidades de financiación internacional y construya una red de contactos.

#### PLUMX

Realice seguimiento a las métricas alternativas de sus productos académicos (descargas, lectores, tweets, entre otros).

Contar con estos perfiles le permitirá disfrutar de los siguientes beneficios:



Fortalecimiento en la visibilidad de su producción.



Acceso a oportunidades de financiación.



Seguimiento exhaustivo e inmediato al impacto generado.



Conexión con pares académicos en sus áreas de interés.



Unificación de su presencia web y de sus citas.

Si tiene inquietudes respecto a sus perfiles académicos, comuníquese con el CRAI en la extensión **3051** o al correo electrónico **investigacion\_crai@urosario.edu.co**.





## La nueva imagen del Repositorio Académico Digital UANL y la implementación de Twitter para la difusión de contenidos

Dr. Carlos Villanueva Valadez<sup>a</sup>  
ORCID:0000-0001-6914-4780

Dirección de Comunicación Institucional UANL, Av. Pedro de Alba, Cd. Universitaria, CP 66451, San Nicolás de los Garza, Nuevo León, México  
[carlos.villanuevava@uanl.mx](mailto:carlos.villanuevava@uanl.mx)

**Resumen:** La Universidad Autónoma de Nuevo León, a través de la Coordinación del Repositorio, ha puesto en marcha el nuevo portal del Repositorio Académico Digital de la UANL, un sitio web que ofrece un diseño atractivo para el usuario, así como una experiencia de interacción con los contenidos y sitios relacionados con el movimiento de acceso abierto a nivel mundial.

La importancia de renovar la imagen del Repositorio se debió a una estrategia de comunicación digital dirigida a estudiantes, docentes, investigadores y sociedad en general, con el objetivo de persuadir con prácticas comunicativas que ayuden a la difusión de la producción académica y científica que es generada en la Universidad.

El uso de redes sociales, concretamente Twitter, proporciona una mayor difusión de los contenidos con los que cuenta el Repositorio, en particular hacia a la comunidad estudiantil, ya que el manejo de estas redes es la práctica comunicativa más utilizada por los estudiantes y es una excelente herramienta para la persuasión que genera una visibilidad de la institución a nivel mundial.

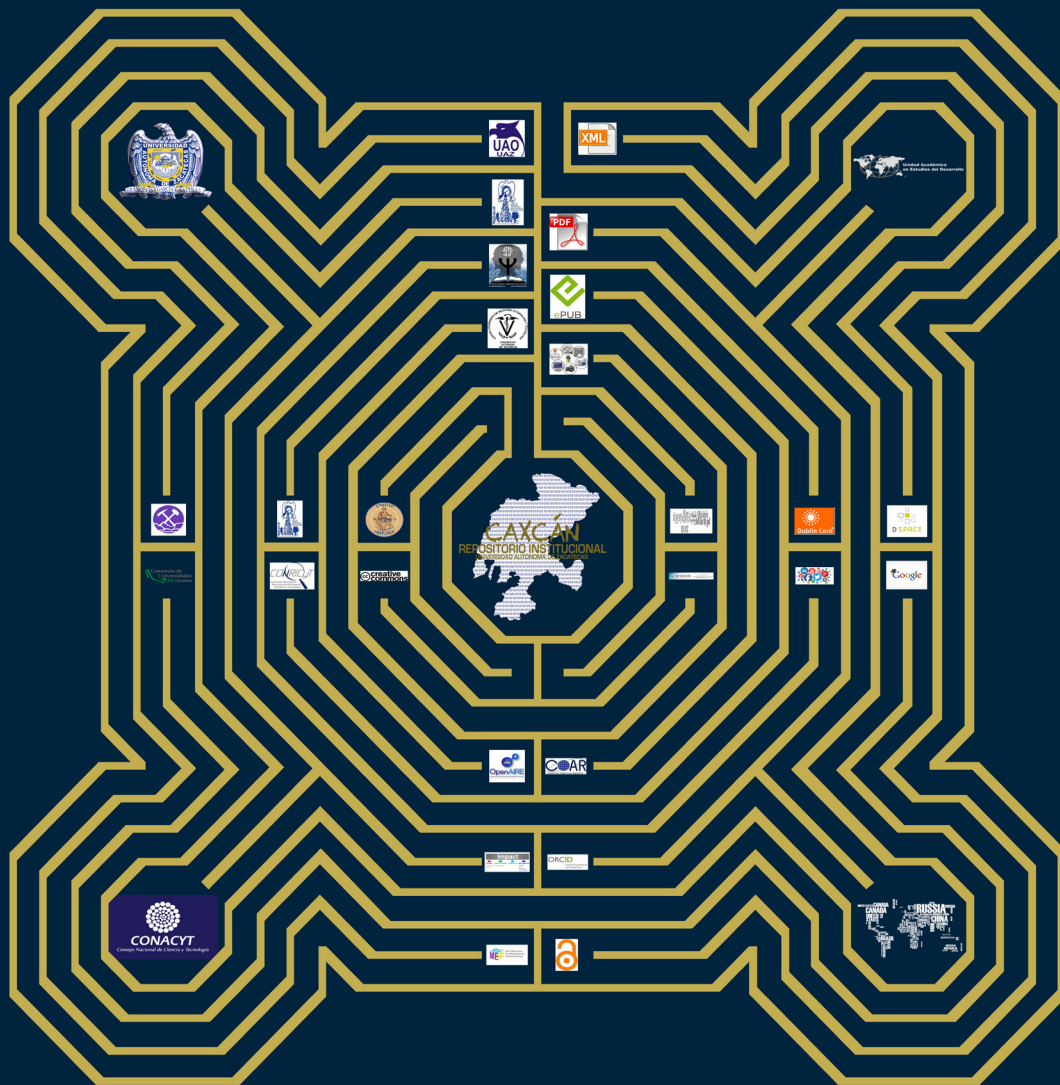




# VI Conferencia internacional BIREDIAL-ISTEC

17-18-19 OCTUBRE 2016

Cynthia Elizabeth Chávez Ceja, Montserrat García Guerrero,  
Elizabeth Gómez Rodríguez, Ana Eleonora Novoa Rivera  
UNIDAD ACADÉMICA EN ESTUDIOS DEL DESARROLLO,  
UNIVERSIDAD AUTÓNOMA DE ZACATECAS



Centro de Documentación e Información  
Unidad Académica en Estudios del Desarrollo  
Universidad Autónoma de Zacatecas

Av. Preparatoria s/n  
Col. Hidraulica  
Zacatecas, Zac.

01 (492) 925 6690 ext. 3517  
documentación@estudiosdeldesarrollo.net





## Publicación en acceso abierto en revistas del núcleo central de la Web of Science en las áreas de Biología y Química

Meilyn Garro Acón, Universidad de Costa Rica. [meilyn.garro@ucr.ac.cr](mailto:meilyn.garro@ucr.ac.cr)  
 Jorge Polanco Cortés, Universidad de Costa Rica. [jorge.polanco@ucr.ac.cr](mailto:jorge.polanco@ucr.ac.cr)

**Tipo de trabajo:** póster  
**Eje temático:** Visibilidad

**Palabras clave:** acceso abierto, publicación científica, Factor de Impacto, cobro por publicar  
**Key words:** open access, scientific publication, Impact Factor, article processing charge

La Universidad de Costa Rica promueve la distribución de su producción académica en acceso abierto. Sin embargo, a nivel institucional esta producción aún es evaluada tomando en cuenta indicadores como el Factor de Impacto y la presencia en ISI y WoS, lo cual genera conflicto entre los investigadores ya que deben sopesar la posibilidad de pagar para publicar en revistas incluidas en esos índices o atenerse a políticas editoriales que los obligan a ceder sus derechos y limitan el acceso a su trabajo. El proceso de selección de una revista se ve afectado por la dificultad para determinar de antemano cuáles publicaciones solicitan pagos, cuáles son los montos de esos pagos y qué políticas tienen las revistas.

**Objetivo:** Verificar las condiciones de apertura de las revistas de acceso abierto en el núcleo central de la Web of Science que declaran ser de acceso abierto.

**Metodología:** Mediante el InCites™ Journal Citation Reports® se hizo un filtrado de las revistas de acceso abierto de Química (chemistry analytical, chemistry applied, chemistry inorganic & nuclear, chemistry medicinal, chemistry multidisciplinary, chemistry organic, chemistry physical) y Biología (biochemistry & molecular biology, biology, cell biology, developmental biology, evolutionary biology, marine & freshwater biology, reproductive biology). Los datos utilizados son de junio de 2016. Se definieron 5 categorías de análisis:

1. Coherencia de políticas de publicación y cobro de APC (Article Processing Charge):

**Charge):** se hizo una clasificación de las revistas con base en la expresión de sus políticas en los sitios web:

- Las que cobran APC y sí indican el monto del APC.
- Las que cobran APC y no indican el monto del APC.
- Las que no indican que no cobran APC.
- Las que sí indican que no cobran APC.

2. **Monto de las APC:** se hizo una lista de precios sumando el monto declarado del APC y otros cobros. Los cobros especificados en cantidad de páginas y figuras se calcularon con base en 12 páginas y tres figuras. Se convirtió todas las monedas a USD.

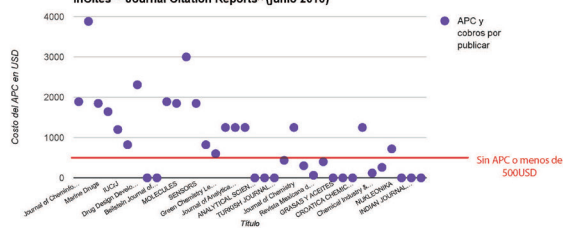
3. **Factor de Impacto de la revista:** se comparó el Factor de Impacto para verificar si las publicaciones de acceso abierto cumplen con los requisitos de entes evaluadores que solicitan a sus autores publicar en revistas con un factor de impacto determinado. Además, se tomó el total de revistas de Biología y Química del núcleo central en la Web of Science para comparar el factor de impacto según tipo de acceso.

4. **Licencia de distribución de la publicación final:** se listaron los permisos para la reutilización por parte de usuarios y autores.

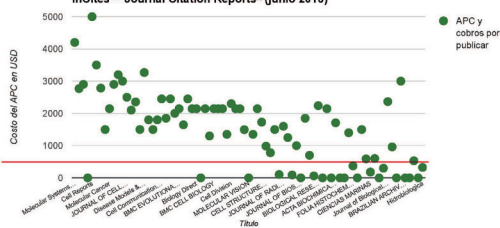
5. **Propiedad de los derechos patrimoniales:** determinación de quién (el autor o la revista) se queda con los derechos patrimoniales de los artículos.

### Resultados

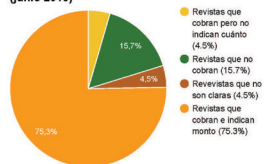
APC y cobros por publicar (USD) en las revistas de Química en InCites™ Journal Citation Reports® (junio 2016)



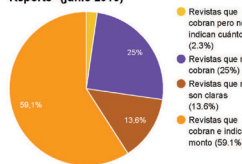
APC y cobros por publicar (USD) en las revistas de Biología en InCites™ Journal Citation Reports® (junio 2016)



Revistas según cobro de APC en el área de Biología en InCites™ Journal Citation Reports® (junio 2016)



Revistas según cobro de APC en el área de Química en InCites™ Journal Citation Reports® (junio 2016)



Se analizaron 89 revistas de Biología y 44 revistas de Química. Un 25% de las revistas de Biología de acceso abierto en la WoS no cobran APC, lo mismo que el 15.7% de Química. En ambas áreas hay revistas de acceso abierto sin cobro de APC con Factor de Impacto  $\geq 1$  ( $IF \geq 1$ ) y un uso extendido de las Licencias Creative Commons (78.5% Biología y 75.6% Química). Los derechos de autor son conservados por los autores en 66.3% de las revistas de Biología y 54.3% de Química.

La mayoría de revistas cobran APC u otras cuotas y hay un porcentaje importante que no tienen políticas claras para los autores, especialmente en Química.

Los APCs oscilan entre los 5000USD y los 60USD en Biología, y los 3883USD y los 64USD en Química. Hay 13 revistas con APCs menor a los 500USD (7 de Biología y 6 de Química).

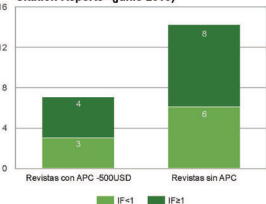
Entre las revistas de Química que no cobran APC hay 2 con  $IF \geq 1$ , y en Biología hay 8. Lo que resulta en un total de 3 revistas con APC igual a 0USD o por debajo de los 500USD y  $IF \geq 1$  de Química y 12 revistas de Biología con esas mismas condiciones.

**Discusión:** Las instituciones de educación superior, al igual que sistemas nacionales de evaluación, emplean sistemas de evaluación para estudiantes, investigadores y docentes que, en ocasiones, incluyen el FI como principal indicador para medir la producción científica (el Doctorado en Ciencias de la Universidad de Costa Rica solicita a sus estudiantes publicar en revistas con  $IF \geq 1$ ).

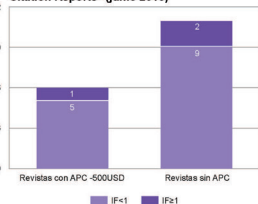
Para los autores, este requisito sumado a las restricciones económicas y a las recientes políticas de acceso abierto limitan la cantidad de revistas en las que pueden publicar. De las 25 revistas de la vía platino (Haschak, 2007) identificadas, solo 10 tienen  $IF \geq 1$ , por lo que la mayor parte de los autores que desean publicar en acceso abierto con  $IF \geq 1$  tendrían que pagar un APC. Por esta razón, es tan importante que, al estipular políticas de evaluación y políticas de acceso abierto, se coordinen las acciones económicas necesarias para poder cumplirlas, ya sea de manera institucional o por medio del ente financiador.

Durante la recolección de datos se evidenció falta de orden para ofrecer la información sobre políticas por parte de las revistas. En general, fue más sencillo encontrar la información en las revistas de Biología que en las de Química, en ambos casos se revisó varias páginas del sitio web para encontrar los datos requeridos y muchas veces fueron confusos.

Revistas con APC <500USD y No APC con  $IF \geq 1$  en el área de Biología en InCites™ Journal Citation Reports® (junio 2016)



Revistas con APC <500USD y No APC con  $IF \geq 1$  en el área de Química en InCites™ Journal Citation Reports® (junio 2016)



#### Referencias

Budapest Open Access Initiative. (2002). Iniciativa de Budapest para el Acceso Abierto. Hungría, Budapest. recuperado en: <http://www.budapestopenaccessinitiative.org/translations/spanish-translation>  
 Cetto, A.M. (2015). Las revistas científicas en América Latina y el Acceso Abierto (AA). Espacio | + D 4 (7).  
 Haschak, P.G. (2007). The 'platinum route' to open access: a case study of E-JASL: The Electronic Journal of Academic and Special Librarianship. Information Research, 12(4) paper 321. [Available at <http://InformationR.net/ir/12-4/paper321.html>]  
 Ruiz-Perez, S. (2010). La edición de revistas científicas en acceso abierto: características editoriales y modelos de negocio en el contexto del Proyecto SOAP (Study of Open Access Publishing). Master thesis, Facultad de Comunicación y Documentación - Universidad de Granada.

Datos completos en [www.kerwa.ucr.ac.cr/handle/10669/28853](http://www.kerwa.ucr.ac.cr/handle/10669/28853)



Todos los derechos reservados Universidad de Costa Rica. Esta obra se comparte bajo una licencia Atribución-NonCommercial-Compartir Igual 3.0 Costa Rica. Para más información visite [ucrdindex.ucr.ac.cr](http://ucrdindex.ucr.ac.cr)

# PATROCINADORES



Thomson Reuters provides professionals with the intelligence, technology and human expertise they need to find trusted answers.

[www.thomsonreuters.com](http://www.thomsonreuters.com)



Unlocking potential with the best learning and research solutions

<http://www.cambridge.org/>



“La librería más antigua de México”:  
LIBRERÍA ESPAÑOLA.

[www.tulibreria.mx](http://www.tulibreria.mx)



Dot.Lib distribuye contenido científico y académico en línea, a cientos de instituciones ubicadas en Latinoamérica, teniendo como socios algunos de los principales editores internacionales.

<http://www.dotlib.com/>



Proporciona soluciones integrales en el manejo de efectivo, digitalizadoras de documentos y equipos de seguridad.

<http://www.seprocesa.com.mx/>



the best of both worlds

Soluciones para bibliotecas que conectan, atraen y evolucionan.

<http://www.bibliotheca.com/>



proveedor líder de servicios de ventas y mercadeo a editoriales académicas y profesionales alrededor del mundo.

[www.accucoms.com/](http://www.accucoms.com/)



Gale is a global leader in education, learning, and research resources online. We have a vast digital archive of research databases, reference books (both eBook and print), and more!

<http://www.gale.com/>



# FOTOGRAFÍAS DEL EVENTO



















# CONCLUSIONES

La VI Conferencia Internacional BIREDIAL-ISTEC 2016 se llevo a cabo por primera vez en México, teniendo como sede las instalaciones del Centro Cultural Universitario Bicentenario de la Universidad Autónoma de San Luis Potosí (UASLP). En esta edición se conto con la participación de 110 asistentes provenientes de Colombia, Costa Rica, Cuba, Argentina, Brasil, México y España. Del interior de la República Mexicana de las siguientes entidades: Ciudad de México, Estado de México, Puebla, Campeche, Guadaluajara, Nuevo León, Colima, Aguascalientes, Guanajuato, Queretaro, Zacatecas, Veracruz, Sonora, Durango, Coahuila y San Luis Potosí.

Como primera actividad social de bienvenida a los visitantes, se organizo con el apoyo de Turismo Municipal de San Luis Potosí, una visita al centro histórico de la ciudad que comprendió la presentación del Recorrido de Leyendas Potosinas con el acompañamiento de la Estudiantina de la UASLP.

El evento de inauguración fue presidido por el Sr. Rector de la UASLP, Maestro en Arquitectura Manuel Fermín Villar Rubio, acompañado por la Dra. Marisa R. De Giusti, Profesora de la Universidad Nacional de la Plata y Representante del Consorcio Iberoamericano para la Educación en Ciencia y Tecnología (ISTEC); el Dr. Miguel Guajardo Mendoza, Director de Análisis Estadístico y Evaluación del Consejo Nacional de Ciencia y Tecnología (CONACYT); y el Dr. Luis del Castillo Mora, Director General del Sistema de Bibliotecas de la UASLP.

Por la noche, se ofreció a los asistentes un cocktail de bienvenida en el lobby principal del Centro Cultural Universitario Bicentenario.

El programa académico estuvo conformado por 3 conferencias magistrales, 4 talleres, un panel con 3 ponentes, tres mesas temáticas con un total de 13 conferencias y una sesión de presentación de posters. En la mesa de Preservación Digital se presentaron 2 trabajos, en la de Interoperabilidad 3 trabajos y en la Visibilidad un total de 8 trabajos.

Cada uno de los trabajos presentados en las diferentes mesas y sesión de posters fueron evaluados y seleccionados por un Comité Científico integrado por investigadores y académicos de instituciones educativas de Argentina, Brasil, Colombia, Costa Rica, España y México. El proceso de aceptación y notificación a autores estuvo a cargo de los de los miembros del Comité organizador internacional de la Conferencia.

En el área de exposiciones se contó con la participación de los siguientes patrocinadores: Librería Española, Gale, Bibliotheca + 3M, Dotlib, Seprocesa, Thomson Reuters, Cambridge University Press y Accucoms.

Para la clausura se llevo a cabo un brindis y evento cultural en el patio del Edificio Central de la UASLP contando con la presentación del Ballet y Mariachi de la UASLP, en el

cual los asistentes disfrutaron de una agradable velada y convivio de despedida.

Para la organización del evento, se integró un comité local conformado por nueve personas de la Biblioteca Virtual Universitaria que durante diez meses estuvieron trabajando arduamente en diferentes tareas, bajo la coordinación general de la Dirección de Biblioteca Virtual Universitaria y la supervisión de la Dirección General del Sistema de Bibliotecas de la UASLP.

Además, se contó con el apoyo y colaboración en diferentes actividades de las siguientes dependencias universitarias: Administración del Centro Cultural Universitario Bicentenario, División de Difusión Cultural, Departamento de Cuotas y Colegiaturas, Departamento de Comunicación Social, Secretaría Administrativa, Departamento de Imagen Institucional, Departamento de Televisión Universitaria, División de Informática y Departamento de Convenios.

Para la difusión y memorias del evento, en los meses previos se hizo el llamado a trabajos y una extensiva campaña de difusión a través del sitio web y página de Facebook de la Conferencia, además de las diferentes páginas web y redes sociales de las instituciones miembros del Comité Internacional. A nivel nacional, se envió la convocatoria a través de la Red Mexicana de Repositorios Institucionales – REMERI, las diferentes redes bibliotecarias como CONPAB, RESBIUC y REBIESNE y los medios informativos de la Corporación de Universidades para el Desarrollo de Internet 2 (CUDI) y del Repositorio Nacional del Consejo Nacional de Ciencia y Tecnología (CONACYT). A nivel local, se realizaron videos promocionales, anuncios publicitarios en medios masivos de comunicación e invitaciones por los diferentes medios institucionales. Durante el desarrollo del evento, se video grabaron todas las sesiones además de toma de fotografías y algunas entrevistas con la prensa y canal de televisión local.

Finalmente, la realización de la VI Conferencia Internacional BIREDIAL-ISTEC 2016 permitió el encuentro e intercambio de conocimientos y experiencias entre investigadores, académicos, profesores, administrativos, técnicos y estudiantes con expertos internacionales en los temas de interés común, estableciendo nuevos contactos y áreas de oportunidad. Los miembros del Comité organizador internacional manifestaron su entera satisfacción por el desarrollo del evento y los resultados obtenidos calificándolo como muy exitoso. Para la UASLP constituyo una oportunidad más para posicionarse a nivel internacional como institución de vanguardia y particularmente para los miembros del comité local, representó un importante reto que asumieron con gran esfuerzo y compromiso, redundando en beneficios para las instituciones y la comunidad académica.



# COMITÉ LOCAL

**Dr. Luis del Castillo Mora**

Director General del Sistema de Bibliotecas  
Supervisor y responsable general del evento

**MTE Rosalina Vázquez Tapia**

Directora de la Biblioteca Virtual Universitaria  
Miembro del Comité Internacional  
Coordinador General del Comité Local

**Lic. Jennifer Escalan.**

Maestra de la Facultad de Ciencias de la Comunicación  
Coordinadora de Logística, proveedores y expositores

**LDG Emmanuel Juárez Castro**

Encargado de Diseño Gráfico y Editorial de la BVU  
Responsable del sitio web, diseño de medios y control de inscritos

**Lic. María Luisa Almendares Marín**

Administradora de la BVU  
Responsable de facturación y control de pagos

**Mtro. Luis Fernando Avila Tobias**

Encargado de Vinculación Académica de la BVU  
Apoyo en el área de difusión, producción audiovisual y fotografía

**Ing. Josué Martínez Vázquez**

Encargado de digitalización de la BVU  
Apoyo a ponentes y conferencistas

**Ing. Edgar Eduardo González Guerrero**

Coordinador Técnica de la Red Universitaria de Videoconferencia, BVU  
Apoyo en la transmisión de ponencias por videoconferencia



**Personal de apoyo en el área de registro:**

Lic. Elizabeth Hernández Peña, BVU

Ing. Claudia Lizett Amaya Varela, BVU

**Personal de apoyo en el área de comunicación:**

Lic. Virginia Carreño Macias

Jefa del Departamento de Comunicación del Sistema de Bibliotecas

Lic. José Adan González Lopéz

Diseñador Gráfico en el Departamento de Comunicación del Sistema de Bibliotecas

**Personal de staff del Centro de Información de Ciencia, Tecnología y Diseño:**

Ing. Rigo José Rigoberto Baltierra Campos,

LB. Jesús Armenta Martínez

LB. Emmanuel Payan Mondragon



Conferencia Internacional  
**BIREDIAL-ISTEC**  
17-18-19 OCTUBRE 2016

- VI CONFERENCIA INTERNACIONAL SOBRE BIBLIOTECAS Y REPOSITORIOS DIGITALES DE AMÉRICA LATINA (BIREDIAL)
- XI SIMPOSIO INTERNACIONAL DE BIBLIOTECAS DIGITALES (SIBD)

**UNIVERSIDAD AUTÓNOMA DE SAN LUIS POTOSÍ**  
**SISTEMA DE BIBLIOTECAS**  
**Biblioteca Virtual Universitaria**  
San Luis Potosí, S.L.P.; México



UNIVERSIDAD DE COSTA RICA



UNIVERSIDAD NACIONAL DE LA PLATA

