

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

Análisis de escenarios y protocolos para integración de repositorios digitales y sitios web institucionales

*de Albuquerque, Pablo, CESGI Comisión de Investigaciones Científicas, pablo@sedici.unlp.edu.ar;
Terrone, Pablo Gabriel, PREBI-SEDICI Universidad Nacional de La Plata,
pabloterrone@sedici.unlp.edu.ar; Villarreal, Gonzalo Luján, PREBI-SEDICI Universidad Nacional de La Plata y CESGI Comisión de Investigaciones Científicas, gonzalo@prebi.unlp.edu.ar; De Giusti, Marisa Raquel, PREBI-SEDICI Universidad Nacional de La Plata y CESGI Comisión de Investigaciones Científicas, marisa.degiusti@sedici.unlp.edu.ar*

Palabras clave

Repositorios digitales, API REST, OAI PMH, Opensearch, Interoperabilidad

Eje temático

Infraestructura tecnológica

Resumen

En este trabajo se analizan los protocolos de interoperabilidad OAI-PMH y OpenSearch, junto a la API REST disponible en DSpace 7, desde el punto de vista de su utilidad a la hora de exponer recursos de repositorios digitales en sitios web institucionales. Se analiza el grado de adopción de cada protocolo, su facilidad de integración con CMS como Wordpress o Drupal, y el nivel de personalización soportado en las consultas y en las respuestas que generan. A partir de este análisis se evalúan distintos escenarios en los cuales es posible utilizar alguno de los protocolos, y se ejemplifica con URL y criterios de filtrado en los casos donde sea posible. La evaluación de estos escenarios sugiere que si bien OAI-PMH resulta adecuado para cosechas masivas, no parece una buena solución cuando se requieren pocos recursos, integrados en CMS, y con criterios de filtrado muy específicos. El protocolo OpenSearch permite una fácil integración con CMS, pero resulta muy limitado por la información que ofrece y la falta de estandarización. Finalmente, la API REST de DSpace 7 ofrece una buena alternativa a OpenSearch, pero su implementación es muy reciente y su integración con CMS más compleja.

Texto completo

Introducción

Los repositorios digitales brindan numerosos mecanismos de interoperabilidad, y con ello buscan fomentar que los recursos alojados por estas plataformas sean integrados dentro de servicios y sistemas de terceros. Esta integración puede realizarse con múltiples fines, como por ejemplo exponer recursos externos para un portal de tesis, un agregador de recursos regional o un blog personal, o como fuente de datos para poblar o agilizar la carga en otro repositorio digital o un CRIS. También sirven para brindar servicios de análisis de producción científica y/o integración de

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

métricas, o incluso para proveer de recursos a sistemas de minería de textos o aprendizaje automático. El grado de detalle con el que se catalogan los recursos en los repositorios digitales, sumado a los estándares de normalización de metadatos utilizados, generan un entorno ideal para que muchos sistemas los aprovechen como fuentes primarias de información.

Las instituciones académicas y científicas cuentan con sitios web institucionales en los cuales describen, con mayor o menor grado de detalle, su organización y estructura interna, los objetivos que persigue y el alcance de sus funciones. Muchas instituciones también utilizan sus sitios web como espacios para la difusión de sus actividades, como noticias, resoluciones, agenda de eventos y logros o premios, y del trabajo que realizan sus miembros, por lo general a través de publicaciones científicas, presentaciones realizadas en congresos, tesis defendidas, patentes registradas, libros publicados e informes de proyectos de investigación. El mantenimiento y actualización periódica del contenido de los sitios web requiere contar con personal capacitado en comunicación institucional y científica; asimismo, a fin de asegurar una cobertura informativa que dé cuenta del trabajo realizado por todas las áreas de la institución, se requiere un compromiso de todos los espacios para comunicar sus actividades y su progreso al área responsable de la gestión del sitio web.

La integración de los repositorios digitales con sitios web institucionales abre las puertas a las instituciones académicas y científicas para ampliar y optimizar la forma de mostrar al mundo la producción de sus propios investigadores. Una mayor exposición podría incidir en el impacto de dicha producción tanto en la comunidad científica como en la sociedad en general. Esta integración puede aplicarse en diferentes contextos, que estarán sujetos a los objetivos que cada institución persigue con su repositorio institucional: tipo de recursos que aloja, estructura de colecciones y comunidades definidas, protocolos de interoperabilidad disponibles, formatos de representación de los recursos, capacidades de consulta y filtrado de datos, entre otros.

Protocolos de interoperabilidad en repositorios digitales

Para poder integrar contenidos de los repositorios digitales con sitios web institucionales, es necesario contar con mecanismos que permitan recuperar recursos desde los repositorios digitales, y mostrarlos en los sitios web institucionales. Dichos sitios podrían estar basados en desarrollos informáticos a medida, aunque por lo general son implementados a través de Sistemas de Gestión de Contenidos (CMS) como Wordpress, Joomla! o Drupal. Esta no es una característica particular de las instituciones académicas, ya que buena parte de los sitios web del mundo se encuentran desarrollados sobre CMS (2023's CMS Market Share Report – Latest Trends and Usage Stats, 2023; Open-source CMS Market Share, 2023), entre los que se destaca Wordpress como el más utilizado a nivel mundial, muy por encima del resto, y los sistemas Drupal y Joomla! dentro de los 10 más utilizados (tabla 1 y gráfico 1)

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

	2022 1 May	2022 1 Jun	2022 1 Jul	2022 1 Aug	2022 1 Sep	2022 1 Oct	2022 1 Nov	2022 1 Dec	2023 1 Jan	2023 1 Feb	2023 1 Mar	2023 1 Apr	2023 1 May	2023 8 May
WordPress	64.2%	64.1%	64.2%	64.3%	64.2%	64.3%	64.2%	64.0%	63.7%	63.5%	63.4%	63.3%	63.3%	63.3%
Shopify	6.5%	6.4%	6.3%	6.2%	6.2%	6.1%	6.1%	6.0%	5.7%	5.5%	5.5%	5.5%	5.6%	5.6%
Wix	3.4%	3.4%	3.4%	3.4%	3.4%	3.4%	3.4%	3.5%	3.6%	3.7%	3.7%	3.7%	3.7%	3.7%
Squarespace	3.0%	3.0%	3.0%	3.0%	3.0%	2.9%	2.9%	2.9%	3.0%	3.0%	3.1%	3.1%	3.1%	3.1%
Joomla	2.5%	2.5%	2.5%	2.5%	2.5%	2.5%	2.4%	2.5%	2.7%	2.7%	2.7%	2.7%	2.7%	2.7%
Drupal	1.9%	1.9%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%	1.8%
Adobe Systems	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.6%	1.7%	1.7%	1.7%	1.7%	1.6%	1.6%
Google Systems	1.4%	1.4%	1.4%	1.4%	1.3%	1.3%	1.3%	1.3%	1.2%	1.2%	1.2%	1.1%	1.1%	1.1%
PrestaShop	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.9%	0.9%	1.0%	1.1%	1.1%	1.1%	1.1%	1.1%
Bitrix	1.2%	1.2%	1.2%	1.2%	1.2%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%	1.1%
Webflow	0.8%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%	0.9%
OpenCart	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%	0.8%

Tabla 1: Market share trends for content management systems. Fuente: https://w3techs.com/technologies/history_overview/content_management

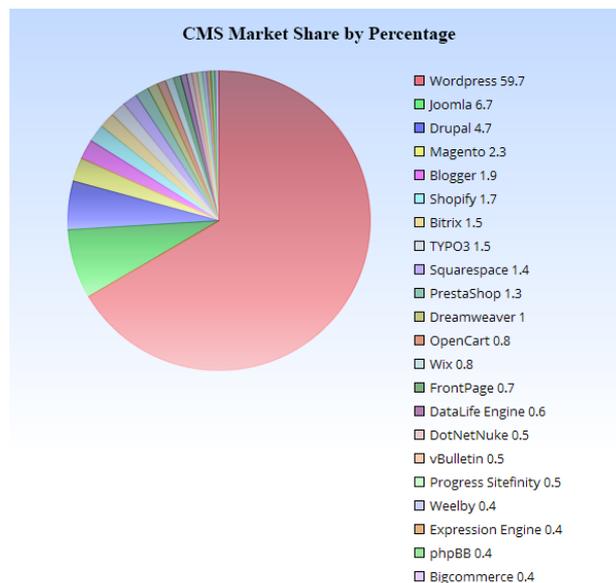


Gráfico 1: Overview of all CMS' with greater than 0.4% market share. Fuente: <https://www.opensourcecms.com/cms-market-share/>

En esta sección se expondrán los protocolos de interoperabilidad más utilizados en el contexto de repositorios digitales y se los analizará como herramientas para exponer recursos en sistemas de terceros. En dicho análisis se definirán criterios como ser:

- forma de recuperación de información
- formato de intercambio de datos
- grado de estandarización del protocolo
- frecuencia de actualización de los datos involucrados

OAI-PMH: Open Archives Initiative - Protocol for Metadata Harvesting

OAI-PMH es un protocolo estándar que permite la recolección y recuperación automatizada de metadatos de recursos digitales desde repositorios. El formato de respuesta utilizado por OAI-PMH

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

es XML, generalmente bajo un esquema de metadatos estandarizado (Dublin Core o Marc por ejemplo), lo que permite una fácil integración con otros sistemas y herramientas.

En cuanto al grado de estandarización, OAI-PMH es un estándar abierto y ampliamente utilizado en el ámbito de la educación y la investigación. Esto significa que los repositorios que implementan este protocolo deben cumplir con ciertas especificaciones y requisitos técnicos, y esto garantiza la interoperabilidad y la compatibilidad con otros sistemas.

En cuanto a la frecuencia de actualización de los recursos expuestos por este medio, esto dependerá de la configuración específica del repositorio y de cómo se implementa OAI-PMH. En general, el protocolo permite a los usuarios realizar solicitudes para obtener metadatos actualizados, lo que significa que la frecuencia de actualización puede ser ajustada según las necesidades del usuario y del repositorio. La recuperación de recursos desde los repositorios digitales puede realizarse siguiendo distintas estrategias. Una de ellas podría ser la implementación de un cosechador OAI-PMH por parte del sitio web, que recupere periódicamente registros desde los repositorios, los almacene localmente y luego muestre subconjuntos de estos recursos en las distintas secciones del sitio web.

El protocolo OAI-PMH permite la definición de diferentes contextos, donde para cada contexto se definen los criterios de filtrado, esto es, qué elementos forman parte del contexto (por ejemplo por tipología documental o colección de origen) y el formato de metadatos con el que se exponen los recursos (por ej. Dublin Core o MARC21). Estos contextos pueden definirse tanto estática como dinámicamente (De Giusti et al., 2014; Mazurek y Werla, 2008). Cabe aquí destacar que el uso de un modelo estandarizado de representación de recursos resulta muy conveniente, ya que simplifica de manera considerable los algoritmos de procesamiento de los registros recuperados. Sin embargo, el uso de este protocolo no parece una solución ideal en caso de necesitar el recurso preservado, ya que el mismo fue pensado para realizar cosechas masivas periódicas, y requeriría que los sitios web institucionales gestionen un gran volumen de recursos en sus bases de datos propias, lo cual excede el propósito de un sitio web institucional tradicional. Asimismo, el uso del protocolo OAI-PMH, al no permitir consultas en vivo, no favorece las actualizaciones automáticas de los contenidos de los sitios web, y para mantener los sitios al día sería necesario realizar cosechas en intervalos de tiempo muy reducidos (diarias, o incluso más de una por día en los casos de instituciones con una gran producción propia) o bien configurar una tarea de indexación periódica en el servidor en función de la cantidad de recursos ingestados en un día.

RSS, OpenSearch y API REST

Para realizar consultas en vivo, una estrategia interesante es el uso del protocolo OpenSearch, basado en RSS/Atom. Este protocolo es soportado por un gran número de sistemas informáticos para repositorios digitales, como por ejemplo DSpace¹, Open Preprint Systems (OPS)², Eprints³ o Digital Commons⁴. Asimismo, muchos repositorios digitales globales o temáticos también brindan

1 El desarrollo de Choique CMS fue discontinuado en el año 2016. Su código fuente se encuentra disponible en el repositorio GitHub <https://github.com/Desarrollo-CeSPI/choique>

2 RSS news feeds for arXiv updates <https://arxiv.org/help/rss>

3 Customizing RSS Output https://wiki.eprints.org/w/Customizing_RSS_Output

4 Featuring Digital Commons Content on External Sites and Platforms

https://bepress.com/reference_guide_dc/featuring-digital-commons-content-external-sites-platforms/

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

servicios de actualizaciones sobre RSS, como por ejemplo Arxiv⁵, PubMed⁶ y bioRxiv⁷. Por el lado de los CMS, un gran número de desarrollos incluyen o permiten incluir fácilmente servicios de recuperación y exposición de RSS de terceros, y de hecho existen experiencias en las que se utilizó el protocolo OpenSearch para exponer contenidos de repositorios digitales en sitios web institucionales (Villarreal et al., 2017), con módulos específicos para Wordpress y Choique-CMS⁸. Por otro lado, a partir de la salida de la versión 7.0 del software DSpace, es también posible realizar consultas en vivo a repositorios con esta herramienta utilizando la nueva API REST de DSpace 7 (DSpace 7 REST API, 2023). A continuación, analizamos algunas ventajas y desventajas de ambas alternativas como soluciones para integrar recursos alojados en repositorios digitales con sitios web institucionales.

Por el lado de OpenSearch, la primera ventaja que podemos observar es que permite consultar y recuperar información utilizando el estándar RSS, basado en el protocolo HTTP, el cual se encuentra incorporado en una amplia variedad de herramientas actuales como lo puede ser Wordpress o Drupal, lo que permite que se puedan incorporar fácilmente las consultas en una gran cantidad de desarrollos. Otra de las ventajas que otorga OpenSearch es la simpleza y estandarización de la estructura de las respuestas que genera este protocolo. Esto hace que procesar la información generada a partir de una consulta sea relativamente sencillo, independientemente del repositorio al que se le realice.

Ya mencionadas las ventajas de este protocolo, es necesario revisar sus puntos débiles. Si bien la respuesta está estandarizada, ofrece una cantidad de metadatos limitada por lo que la información que es posible extraer de cada ítem se encuentra bastante acotada. Además, el protocolo RSS genera las respuestas en formato XML, mientras que un número cada vez mayor de sistemas están comenzando a utilizar JSON como formato de transporte de datos, gracias a su facilidad de procesamiento y menor tamaño requerido para representar información estructurada.

También es necesario analizar la construcción de consultas con OpenSearch. Para empezar, el formato de las consultas pueden variar en cada repositorio, ya que las mismas poseen muchos parámetros configurables. Por ejemplo, para consultar por autor en el repositorio SEDICI usamos la palabra clave *author*, pero en el repositorio institucional de CONICET se utiliza la *keyword dc.contributor.author*. Sumado a lo anterior, los parámetros OpenSearch disponibles en distintos repositorios digitales pueden encontrarse muy limitados, lo que en ocasiones no permite obtener el conjunto de datos deseado. Por ejemplo, si un repositorio organiza sus recursos en colecciones vinculadas a Centros de Investigación y no a tipologías documentales, y no permite filtrados por tipologías mediante OpenSearch, será muy complejo recuperar todos los recursos de un tipo particular (por ejemplo libros) realizados en un centro de investigación (en términos de OpenSearch, filtrando por el *scope*).

En resumen, si bien el protocolo OpenSearch es fácil de utilizar, gracias a estar basado en RSS y a la estandarización y simpleza en sus metadatos, posee algunas desventajas a la hora de consultar distintos repositorios, ya que requiere configuraciones particulares para cada caso, y posee poca

5 Configure RSS feeds - LYRASIS Wiki

<https://wiki.lyrasis.org/display/DSPACE/Configure+RSS+feeds>

6 Add support for RSS feeds to OPS <https://github.com/pkp/pkp-lib/issues/7623>

7 Create RSS Feeds in PubMed

https://www.nlm.nih.gov/pubs/techbull/ja20/ja20_pubmed_updated.html

8 bioRxiv Alerts/RSS <https://www.biorxiv.org/alertsrss>

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

flexibilidad para parametrizar las consultas. Asimismo, este protocolo limita de manera considerable la información que es posible obtener de las fuentes.

Como alternativa al protocolo OpenSearch, es importante analizar las ventajas y desventajas del uso de la API REST incorporada en la versión de DSpace 7 (DSpace 7 REST Contract, 2023). Entre los principales puntos positivos se destaca la estandarización del formato de los *requests* HTTP para construir las consultas; al tratarse de una API REST, los *endpoints* a consultar y filtros a utilizar serán siempre los mismos entre los distintos repositorios digitales, por lo que prácticamente no hace falta realizar configuraciones a la hora de construir las consultas. Además, esta API REST permite una gran cantidad de filtros y parámetros, lo cual brinda flexibilidad a la hora de obtener un conjunto de datos acotado y preciso de acuerdo a la información buscada. Por último, las respuestas que ofrece este mecanismo se encuentran en formato JSON, y poseen mucha más información sobre cada ítem en comparación con la que otorga las respuestas de consultas por OpenSearch.

En cuanto a los puntos negativos de la API REST, podemos destacar que las respuestas no poseen campos estandarizados por lo que procesarlas puede requerir configuración individual para cada repositorio. Además, en el contexto de análisis de este trabajo, los sistemas de gestión de contenidos no suelen estar integrados con API REST del mismo modo que lo hacen con OpenSearch, lo que requerirá esfuerzos adicionales de desarrollo a la hora utilizar este protocolo para recuperar recursos desde repositorios digitales y procesar las respuestas en formatos JSON.

Escenarios

A continuación, se describirán algunos escenarios posibles en los que es posible aprovechar las capacidades de interoperabilidad para integrar repositorios digitales con sitios web institucionales. Si bien quedarán muchos escenarios por imaginar, el listado propuesto sirve para indagar sobre distintas formas de integración, y revisar requerimientos y limitaciones de las herramientas y tecnologías analizadas. Para cada escenario se describe su utilidad, y se incluyen ejemplos concretos sobre su uso o aplicabilidad en contextos reales. Asimismo, algunos escenarios incluyen también requerimientos técnicos particulares que deben cumplirse, o usos alternativos de acuerdo a las facilidades técnicas disponibles en cada contexto.

Escenario 1: exposición de reglamentaciones

Las instituciones académicas pueden mostrar en sus sitios web institucionales las reglamentaciones, resoluciones y ordenanzas publicadas y almacenadas en su repositorio institucional. Para ello, el protocolo de interoperabilidad deberá permitir recuperar este tipo de recursos, ya sea a partir de su tipología documental o de la colección a la que pertenece. Esta información puede ser expuesta agrupada por año y por categoría, y ordenada por fecha de publicación. Por ejemplo, el siguiente enlace permite acceder a las resoluciones aprobadas por la Universidad Nacional de La Plata durante el período 2010-2019 a través de su repositorio SEDICI:

http://sedici.unlp.edu.ar/handle/10915/18180/discover?filtertype=dateIssued&filter_relational_operator>equals&filter=%5B2010+TO+2019%5D

Y utilizando el protocolo OpenSearch, puede recuperarse a través de la siguiente URL de consulta:
<http://sedici.unlp.edu.ar/open-search/discover?>

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

[rpp=10&format=atom&sort_by=3&order=desc&scope=10915/18180&start=0&query=\(dateIssued:%222017%22%20or%20dateIssued:%222018%22%20or%20dateIssued:%222019%22\)](http://rpp=10&format=atom&sort_by=3&order=desc&scope=10915/18180&start=0&query=(dateIssued:%222017%22%20or%20dateIssued:%222018%22%20or%20dateIssued:%222019%22))

De este modo, cada vez que algún área de la institución aprueba una nueva reglamentación y realice el depósito de la misma en el repositorio institucional, y que desde este espacio se completen los procesos de catalogación, descripción y preservación, se mostrará de inmediato en el sitio web de la institución. Esto resulta de gran utilidad para exponer en todo momento la información más actualizada vinculada a este tipo de documentos, incluso cuando su generación se realiza de manera descentralizada.

Escenario 2: producción científica de centros de investigación

Los centros de investigación pueden exponer en sus sitios web la producción de artículos científicos, trabajos en congresos y proyectos de investigación. Por ejemplo, el Centro de Geología Aplicada, Agua y Medioambiente (CGAMA) de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) y de la Universidad Nacional del Sur (UNS) expone en su sitio web todos los artículos publicados por sus investigadores y alojados en el repositorio institucional CIC DIGITAL. En el siguiente enlace puede observarse esta integración:

<https://cgama.cic.gba.gob.ar/publicaciones/articulos>

Para obtener estos recursos, el sitio web del CGAMA puede realizar la siguiente consulta al repositorio CIC-DIGITAL por medio de la API REST de Dspace 7:

<https://host170.sedici.unlp.edu.ar/server/api/discover/search/objects?scope=c0c159b9-a6aa-4548-b326-048d7169d757>

Escenario 3: lecturas recomendadas

En el contexto de entornos de enseñanza virtuales o LMS, así como también en sitios web de cátedras y departamentos, los docentes pueden incluir listados de recursos de la propia institución que los alumnos podrían aprovechar para enriquecer sus conocimientos y fortalecer los conceptos vistos durante las clases. Aquellas instituciones que posean Repositorios de Objetos de Aprendizaje pueden también aprovechar estas plataformas para incorporar recursos en sus propios LMS (Kalou et al. 2008). En este escenario, resulta esencial contar — del lado del repositorio institucional — con capacidades de filtrado de documentos por área temática, a fin de ofrecer a los alumnos recursos académicos elaborados por cualquier investigador, independientemente del departamento o centro de investigación en el que realice sus actividades. Por ejemplo, el repositorio SEDICI permite consultas a través de OpenSearch filtrando por subject, como puede verse en este ejemplo:

[http://sedici.unlp.edu.ar/open-search/discover?rpp=10&start=0&query=\(subject:"economía"\)](http://sedici.unlp.edu.ar/open-search/discover?rpp=10&start=0&query=(subject:)

En el caso de repositorios digitales que funcionen sobre DSpace 7, la consulta puede resolverse consultando directamente los datos expuestos a través del módulo de búsqueda "discovery". En el siguiente ejemplo se recuperan los recursos cuya palabras clave sea "Patrimonio" en el repositorio CIC-DIGITAL:

<https://host170.sedici.unlp.edu.ar/server/api/discover/search/objects?f.keyword=patrimonio,equals>

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

Escenario 4: producción de alumnos de un curso o carrera

Los responsables del dictado de cursos y/o carreras pueden utilizar los repositorios de su institución para alojar y difundir trabajos de cursos, trabajos finales integradores de carreras, tesinas de grado y tesis de postgrado. Al hacer esto, podrán exponer en sus sitios web los trabajos realizados por sus propios alumnos, lo cual puede servir como ejemplo a otros alumnos para desarrollar sus propios trabajos y también como base para continuar y extender trabajos previos. Esto resulta en particular interesante para las áreas dedicadas a coordinar carreras de postgrado, ya que podrán mostrar a sus alumnos actuales, así como también a personas interesadas en realizar alguna carrera, las tesis de doctorado y maestría que han realizado otros alumnos. Esta funcionalidad también puede resultar útil para organismos financiadores de becas doctorales, ya que podrían mostrar en sus propios sitios web las tesis ya aprobadas por sus becarios. Por ejemplo, el siguiente enlace permite acceder a las tesis doctorales realizadas por alumnos de la Facultad de Informática de la UNLP desde el repositorio SEDICI:

http://sedici.unlp.edu.ar/handle/10915/36/discover?filtertype=type&filter_relational_operator>equals&filter=Tesis+de+doctorado

y la siguiente URL permite recuperar esta información por OpenSearch:

[http://sedici.unlp.edu.ar/open-search/discover?format=atom&scope=10915/36&query=\(sedici.subtype:%22Tesis%20de%20doctorado%22\)%22](http://sedici.unlp.edu.ar/open-search/discover?format=atom&scope=10915/36&query=(sedici.subtype:%22Tesis%20de%20doctorado%22)%22)

Si se intenta obtener esta información desde el repositorio CONICET DIGITAL, no es posible obtener tesis doctorales de la misma Facultad, pero sí es posible obtener todas las tesis doctorales a través de la siguiente URL:

https://ri.conicet.gov.ar/discover?filtertype_0=type&filter_relational_operator_0=contains&filter_0=thesis&filtertype=dateIssued&filter_relational_operator>equals&filter=2019

Sin embargo, dado que la implementación de OpenSearch de CONICET DIGITAL no soporta filtrado por tipo de documento, no es posible obtener el mismo conjunto de recursos mediante este protocolo. En estos casos, la falta de estandarización a la hora de implementar interfaces OpenSearch en los repositorios digitales presenta un problema para brindar un mismo servicio de recuperación de datos desde múltiples repositorios.

Escenario 5: gobierno abierto y transparencia

Las instituciones financiadas con fondos públicos deben promover la transparencia en todas sus acciones y brindar a la población general información sobre contrataciones, licitaciones, inversiones y compras en general. En el caso de Argentina, se fomentan y generan espacios de interacción entre la sociedad y el Estado para articular la colaboración y participación ciudadana en las distintas etapas del ciclo de las políticas públicas, y esto ha llevado a la implementación de distintos planes vinculados a Datos Abiertos (Datos Abiertos, 2023) y a gobierno abierto (Quinto Plan de Acción Nacional de Gobierno Abierto (2022-2024), 2023). Dichos planes involucran también a las instituciones del Sistema Nacional de Ciencia y Tecnología (SNCT), que pueden aprovechar sus repositorios digitales para recuperar los documentos vinculados a concursos de incorporación de personal o de cambios de categorías, reportes de licitaciones, informes de compras y contrataciones

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

a terceros. Por ejemplo, el repositorio SEDICI posee una colección dedicada Documentos Institucionales de la UNLP (<http://sedici.unlp.edu.ar/handle/10915/71561>), desde la cual es posible acceder a sub-colecciones y comunidades dedicadas a auditorias, planes estratégicos, licitaciones y concursos. En este sentido, es posible recuperar las licitaciones a través de la interfaz OpenSearch por medio de la siguiente URL:

<http://sedici.unlp.edu.ar/open-search/discover?rpp=10&start=0&scope=10915/69217>

Escenario 6: perfiles de autores

Un uso muy interesante de las facilidades de interoperabilidad entre repositorios digitales y sitios web institucionales es la posibilidad de generar perfiles institucionales de investigadores, docentes, becarios y autoridades en general, que expongan toda su producción científica y académica y que se actualicen de manera automática. En este caso, cada perfil puede incluso mostrar la producción alojada en diferentes repositorios, tanto institucionales –lo que resulta muy conveniente para autores de múltiple dependencia– como también globales o temáticos, como por ejemplo arXiv o PubMed.

Por ejemplo, el sitio web del Instituto Argentino de Radioastronomía posee páginas dedicadas a sus miembros, y en dichas páginas puede verse la producción científica recuperada directamente desde el repositorio CONICET DIGITAL. Esto puede observarse en la página del Dr. Gustavo Romero, disponible en este enlace:

<https://www.iar.unlp.edu.ar/institucional/personal/dr-gustavo-e-romero/>

Para casos de autores con múltiples dependencias, puede tomarse como ejemplo los perfiles de investigadores del Centro de Servicios en Gestión de Información (CESGI) de la CIC. Aquí puede observarse el perfil de la Dra. Marisa R. De Giusti, que incluye publicaciones recuperadas desde tres repositorios institucionales: SEDICI (UNLP) CIC-DIGITAL (CIC) y CONICET-DIGITAL (CONICET):

<https://prebi.unlp.edu.ar/personal/marisa-raquel-de-giusti/>

Escenario 7: sistemas de agregación de recursos académicos

Los sistemas de agregación de recursos académicos tienen como objetivo recopilar items desde distintos repositorios y agruparla en un mismo conjunto, son sistemas que manejan grandes volúmenes de información y la misma suele ser actualizada periódicamente; un ejemplo de esto es Sistema Nacional de Repositorios Digitales Argentino, el cual recopila recursos desde múltiples repositorios institucionales de Argentina, lo que permite al Ministerio de Ciencia, Tecnología e Innovación (MINCYT) mostrar desde un único espacio virtual la producción académica nacional al mundo. Asimismo, los recursos recuperados por el SNRD pueden ser cosechados por LA Referencia utilizando el mismo protocolo e integrándolos con recursos provenientes de otros países, lo que le permite generar servicios de valor agregado sobre un conjunto de recursos mucho más amplio. En este tipo de sistemas, la utilización del protocolo OAI-PMH resulta de gran utilidad, especialmente gracias al grado de estandarización de metadatos, las directrices y acuerdos generados entre las partes, y la facilidad de ejecutar cosechas masivas de recursos.

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

Análisis y conclusiones

Los repositorios digitales proveen diferentes mecanismos para recuperar contenidos y exponerlos en sitios web institucionales. Entre estos mecanismos se destaca el protocolo OAI PMH como protocolo adecuado para la recuperación de grandes volúmenes de información de manera periódica, y los protocolos OpenSearch y API REST como protocolos para la recuperación de volúmenes de información en vivo y más acotados utilizando filtros. Si bien el protocolo OAI PMH posee una amplia madurez y se encuentra muy afianzado en el ámbito de los repositorios digitales, su integración en sitios web institucionales resulta algo compleja debido a la poca integración nativa que por lo general ofrecen los CMS con este protocolo, y parece más adecuado para situaciones donde sea necesario mostrar una gran cantidad de recursos y generar servicios como búsqueda y filtrado sobre los mismos. Por otro lado, alternativas para consultas en vivo como el protocolo OpenSearch y la API REST de Dspace 7 parecen convenientes a la hora de recuperar conjuntos de recursos más acotados, basados en filtros que pueden ser más dinámicos y combinables. En este sentido, el protocolo OpenSearch resulta de momento más simple de integrar con sitios web institucionales implementados con gestores de contenidos como Wordpress o Drupal, gracias a estar basado en el estándar RSS y a que estos sistemas ya incorporan o permiten incorporar muy fácilmente soporte para la recuperación y exposición de información a través de este protocolo; sin embargo, el protocolo RSS presenta importantes limitaciones en cuanto a la cantidad y a la estructura de información que permite recuperar y, por lo tanto, en cuanto al tipo de servicios que es posible brindar a partir del mismo. Por el otro lado, el uso de la API REST ofrece una mayor flexibilidad a la hora de construir consultas y permite recuperar una mayor cantidad de información de cada recurso; sin embargo, al tratarse de un desarrollo relativamente nuevo, aún no ha sido adoptado por un gran número de repositorios digitales y quizás su adopción masiva demore algunos años más. Asimismo, su integración con los CMS más utilizados hoy en día resulta algo más compleja comparada con OpenSearch, ya que requiere implementar módulos ad-hoc para generar y ejecutar los *requests* a la API REST, y procesar los resultados en formato JSON antes de presentarlos al usuario final.

Bibliografía

1. 2023's CMS Market Share Report – Latest Trends and Usage Stats (2023). *WPbeginner.com* . Recuperado el 8 de mayo de 2023 de <https://www.wpbeginner.com/research/cms-market-share-report-latest-trends-and-usage-stats/>
2. Datos Abiertos (2023). *Argentina.gob.ar*. Recuperado el 8 de mayo de 2023 de <https://www.argentina.gob.ar/jefatura/innovacion-publica/servicios-y-pais-digital/gobierno-abierto/datos-abiertos> .
3. De Giusti, M. R., Adorno, F. G., Lira, A. J. (2014). Repositorios DSpace con múltiples contextos OAI-PMH. En *IV Conferência Internacional sobre Bibliotecas e Repositórios Digitais (BIREDIAL) y IX Simpósio Internacional de Bibliotecas Digitais (SIBD) (Porto Alegre, Brasil, 2014)*. <http://sedici.unlp.edu.ar/handle/10915/44572>

BIREDIAL-ISTEC 2023

XII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales
Del 18 al 20 de octubre de 2023

4. *Dspace 7 REST Contract* (2023). *GitHub.com*. Recuperado el 8 de mayo de 2023 de <https://github.com/DSpace/RestContract/>
5. *Dspace 7 REST API* (2023). *Lyrasis.org*. Recuperado el 8 de mayo de 2023 de <https://wiki.lyrasis.org/display/DSPACE/DSpace+7+REST+API>
6. Kalou, A.K., Koutsomitropoulos, D.A., Solomou, G.D., Botsios, S.D. (2015). *Metadata Interoperability and Ingestion of Learning Resources into a Modern LMS*. In: Garoufallou, E., Hartley, R., Gaitanou, P. (eds) *Metadata and Semantics Research. MTSR 2015. Communications in Computer and Information Science*, vol 544. Springer, Cham. https://doi.org/10.1007/978-3-319-24129-6_15
7. Mazurek, C., & Werla, M. (2008). Extending OAI-PMH protocol with dynamic sets definitions using CQL language. En *Conference proceedings of IADIS Information Systems, Algarve, Portugal (April 2008)*. IADIS Publications
8. Open-source CMS Market Share (2023). *OpenSource CMS*. Recuperado el 8 de mayo de 2023 de <https://www.opensourcecms.com/cms-market-share/>
9. Quinto Plan de Acción Nacional de Gobierno Abierto (2022-2024) (2023). *Argentina.gob.ar*. Recuperado el 8 de mayo de 2023 en <https://www.argentina.gob.ar/jefatura/innovacion-publica/servicios-y-pais-digital/gobierno-abierto/quinto-plan-de-accion-nacional>
10. Villarreal, G. L., Salamone Lacunza, P., Vila, M. M., De Giusti, M. R., & Manzur, E. (2017). *A Simple Method for Exposing Repository Content on Institutional Websites*. *Open Repositories*. En *Open Repositories 2017* (Queensland, Australia, 2017). <http://sedici.unlp.edu.ar/handle/10915/60507>