

# LOS NUEVOS ROLES DEL REPOSITORIO INSTITUCIONAL

PALABRAS CLAVE:

COMUNICACIÓN CIENTÍFICA - ACCESO ABIERTO - REPOSITORIOS DIGITALES

Por Marisa Raquel De Giusti

**RESUMEN:** Este trabajo tiene como objetivo principal compartir la experiencia de la gestión de la producción académica y científica que se realiza en un repositorio institucional, así como los nuevos desafíos que han surgido como producto de nuevas expectativas de los usuarios de repositorios, y por las posibilidades que brindan las nuevas tecnologías y estándares aún no incluidos en los repositorios actuales. Reconoce el cambio fundamental que se ha producido en la comunicación científica debido a nuevas prácticas dadas por el advenimiento de la web 2.0 y las redes sociales, así como las novedades de las denominadas Tecnologías de la Información y la Comunicación (TIC) que han transformado el espacio de posibilidades para la creación y difusión del conocimiento.

El artículo recorre las nuevas modalidades en el contexto amplio de lo que se denomina acceso abierto a la ciencia y, a la luz de sus posibilidades, pretende dar a conocer a una comunidad amplia que se encuentra comprometida en el proceso de crear repositorios. Asimismo, analiza el futuro para los repositorios que propone la Confederación Internacional de Repositorios de Acceso Abierto (COAR) así como otras propuestas e investigaciones relacionadas que buscan roles más activos para quienes intervienen en la generación de conocimientos académicos y científicos.

## INTRODUCCIÓN

Desde sus albores, la iniciativa de acceso abierto - ratificada en las famosas declaraciones de Budapest<sup>1</sup>, Bethesda<sup>2</sup> y Berlín<sup>3</sup>-, propuso la difusión libre y gratuita del conocimiento ofreciendo dos vías alternativas para plasmar la distribución de los contenidos científicos: las revistas de acceso abierto y los repositorios digitales. Sabiamente, la declaración de Budapest hablaba de la convergencia de la vieja tradición de los científicos de compartir desinteresadamente sus obras y el advenimiento de una nueva tecnología que permitía nuevas modalidades haciendo uso de internet. El antiguo sistema de comunicación científica detentado por las editoriales, que cada día son menos y abarcan más títulos, representa la modalidad tradicional, comercial y, a la vez, el circuito de evaluación de la propia ciencia donde las esferas de producción y difusión del conocimiento estaban aisladas.

El acceso abierto y las nuevas tecnologías de crecimiento exponencial han abierto nuevas posibilidades para democratizar el conocimiento y acelerar la difusión de los contenidos rompiendo la estancamiento del circuito tradicional<sup>4</sup> y han empujado a las grandes editoriales a transformarse, sin por ello perder la posibilidad de rédito. Desde el acceso abierto se ve con

preocupación que a los grandes costos de suscripción, los editores más importantes hayan propuesto un acceso abierto (vía dorada) para revistas de corriente principal con altísimos costos de APCs<sup>5</sup>, obteniendo entonces una nueva fuente de ganancias a costa de los pagos por parte de autores e instituciones. Incluso en Europa se plantea como posibilidad para obtener un acceso abierto de manera más rápida, la propuesta discutida en el marco de Horizonte 2020 que se denomina coloquialmente “*the flip*” (darse vuelta) y pagar por APCs de manera general a las editoriales en lugar de pagar suscripciones<sup>6</sup>.

Muchas organizaciones como UNESCO y COAR se han expedido rechazando drásticamente esta propuesta que institucionalizaría un circuito de pago diferente, pero no por ello menos perverso porque sigue restringiendo el acceso y acentúa las

diferencias, incluso en aspectos importantes como la participación en las selecciones temáticas, privilegiando temas de gran importancia en el ámbito de las naciones más ricas frente a temáticas regionales o locales de importancia social, así como la generalización de la lengua oficial de la comunicación científica.

## PROBLEMAS ADICIONALES EN EL CIRCUITO DE COMUNICACIÓN CIENTÍFICA

La publicación de resultados de investigación tiene siempre dos motivaciones: la primera, compartir avances y conocimientos que sirvan para nuevos avances; la segunda está vinculada a los ascensos y reconocimientos, premios o castigos para los investigadores que suceden tras los procesos de evaluación.

Los países latinoamericanos, y particularmente la Argentina, utilizan un sistema de evaluación que puede llamarse “simplificado”, que consiste en el conteo de publicaciones en un período determinado y en la localización de las revistas en las que se ha publicado y muy particularmente su posicionamiento en un número restringido de índices que las ubican en los denominados cuartiles. El cuartil es un indicador que sirve para evaluar la importancia relativa de una revista dentro del total de revistas de su área. Es una medida de posición de una revista en relación con todas las de su área, que divide en cuatro partes iguales un listado de revistas ordenadas de mayor a menor según el índice de impacto.

En definitiva, la utilización del factor de impacto como modo de evaluar una carrera desconoce entre otras cuestiones, los orígenes y las motivaciones de la pro-

1. Budapest Open Access Initiative | Spanish Translation. (<http://www.budapestopenaccessinitiative.org/translations/spanish-translation>). Consulta 17/11/2017.

2. Bethesda Statement on Open Access Publishing. (<http://legacy.earlham.edu/~peters/fos/bethesda.htm>). Consulta 11/01/2018.

3. Berlin Declaration. (<https://openaccess.mpg.de/Berlin-Declaration>). Consulta 11/01/2018.

4. Orduna-Malea, E., & Delgado López-Cózar, E. (2017). Taller: Research Analytics: la monitorización del impacto académico de un investigador. VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIRELAL-ISTEC'17) y XII Simposio Internacional de Biblioteca Digitales (SIBD'17) (La Plata, 2017). (<http://hdl.handle.net/10915/63599>). Consulta 16/03/2018.

5. Shamash, Katie. "Article processing charges (APCs) and subscriptions - Monitoring open access costs". *Jisc*, 2016, p. 6-15.

6. OA 2020: Open Access 2020 international initiative pushing towards Open Access publishing | Agricultural Information Management Standards (AIMS).

la existencia de ese indicador, la divergencia de modos de publicación según las áreas del conocimiento y la diferencia en cantidad de publicaciones periódicas en las distintas áreas del saber, ocasionada, precisamente, por los diferentes modos de circulación del saber. Un amplio espectro de científicos referentes ha marcado tales sesgos y críticas concretas; para un conocimiento más detallado puede verse el curso impartido por Isabel Bernal<sup>7</sup>.

Sale<sup>8</sup> habla de “los dos sombreros” como figura que representa los distintos comportamientos de los propios investigadores: a la hora de actuar como lectores, pretenden acceso libre y gratuito; en cambio como autores, se inclinan por revistas de pago por suscripción o con altas APCs para lograr una evaluación positiva de sus carreras. Se trata finalmente, en palabras de Albagli y Maciel<sup>9</sup>, de la “socialización del conocimiento o su privatización” pero el sistema de evaluación perpetúa y acentúa el circuito de comercialización en desmedro de un contexto, el de los países a favor del acceso abierto, plasmado en una nueva legislación, como por ejemplo en Argentina la Ley 26.899 de Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos<sup>10</sup> y la Ley de Derecho de Acceso a la Información Pública<sup>11</sup>. Lo precedente, sin lugar a dudas confun-

## El acceso abierto y las nuevas tecnologías de crecimiento exponencial han abierto nuevas posibilidades para democratizar el conocimiento y acelerar la difusión de los contenidos rompiendo la estanqueidad del circuito tradicional.

de al investigador que tradicionalmente no se ocupa de analizar los derechos que cede sobre sus obras o las obligaciones que surgen de la legislación nueva. Estos conflictos se agravan porque los modos de implementación del acceso abierto no son sencillos y la implementación que acompaña a la legislación, muchas veces no es clara en los materiales que alcanza, las obligaciones contraídas por los investigadores y las penalizaciones en caso de no cumplimiento, particularmente en la experiencia latinoamericana.

### CONTEXTO INTERNACIONAL Y POLÍTICAS DE ACCESO ABIERTO. LAS REVISTAS INDEXADAS

Como es de público conocimiento, el acceso abierto se ha plasmado a lo largo del tiempo en dos vías que han tenido distinto crecimiento y distintos niveles de

adopción: la vía dorada de las revistas de lectura libre y gratuita lo que, como se dijo, no siempre significa que todo el circuito sea gratuito, y los repositorios de acceso abierto que idealmente son los espacios institucionales donde los autores autoarchivan sus producciones. La vía dorada se ha complejizado y dentro de lo que se puede denominar la tradición de la ciencia resulta difícil imponer revistas (por fuera de las editoriales comerciales) de acceso abierto como competidoras reales de un pequeño, cuando menos acotado, número de revistas de las grandes editoriales. Por ejemplo, la base de datos Scopus de Elsevier en su listado 2017 reporta 23.496 títulos activos mientras que el directorio extendido de publicaciones periódicas de Ulrich cuenta con aproximadamente 63.000 títulos. Cabe pregun-

7. Bernal, Isabel. Catalogación con DublinCore, gestión de derechos de autor y permisos de editores. Medida, visibilidad e impacto de la producción científica. ([http://digital.csic.es/bitstream/10261/36464/1/Curso\\_Impacto\\_DigitalCSIC.pdf](http://digital.csic.es/bitstream/10261/36464/1/Curso_Impacto_DigitalCSIC.pdf)). Consulta 16/03/2018.

8. Sale, A. "Researchers and institutional repositories". En N. Jacobs (Ed.), *Open Access: Key Strategic, Technical and Economic Aspects* (pp. 87-100). Oxford: Chandos Publishing, 2006.

9. Albagli, S. & Maciel, M. L. *Informação, conhecimento e poder: mudança tecnológica e inovação social*. Garamond, 2011. (<http://livroaberto.ibict.br/handle/123456789/1062>). Consulta 16/03/2018.

10. Argentina. Ley 26899: Creación de Repositorios Digitales Institucionales de Acceso Abierto, Propios o Compartidos, 2013. (<https://www.boletinoficial.gob.ar/#!DetalleNorma/98996/20131209>). Consulta 16/03/2018.

11. Argentina. Ley 27.275 de Derecho de Acceso a la Información Pública, 2016. (<https://www.boletinoficial.gob.ar/#!DetalleNorma/151503/20160929>). Consulta 16/03/2018.

12. Mongeon, P., & Paul-Hus, A. "The journal coverage of Web of Science and Scopus: a comparative analysis". *Scientometrics*, 2016, vol. 106, N° 1, p. 213-228. (<https://doi.org/10.1007/s11192-015-1765-5>). Consulta 16/03/2018.

13. Cetina, K. K. "Epistemic Cultures: Forms of Reason in Science". *History of Political Economy*, 1991, vol. 23, no. 1, p. 105-122. (<https://doi.org/10.1215/00182702-23-1-105>). Consulta 16/03/2018.

14. Larivière, É. A. D. C. Y. G. V. "Comparing bibliometric statistics obtained from the Web of Science and Scopus". *Science-Metrix*, 2009, vol. 60, N° 7, p. 1320-1326.

15. DOAJ (Directory of Open Access Journals). (<https://doaj.org>). Consulta 12/01/2018.

16. Open Access Journals in the 2015 Journal Citation Process. 2017, marzo 16. (<https://clarivate.com/blog/science-research-connect/open-access-journals-2015-journal-citation-process/>). Consulta 12/01/2018.

17. Björk, B.-C., & Solomon, D. Article processing charges in OA journals: relationship between price and quality. *Scientometrics*, 2015, vol. 103, N° 2, p. 373-385. (<https://doi.org/10.1007/s11192-015-1556-z>). Consulta 16/03/2018.

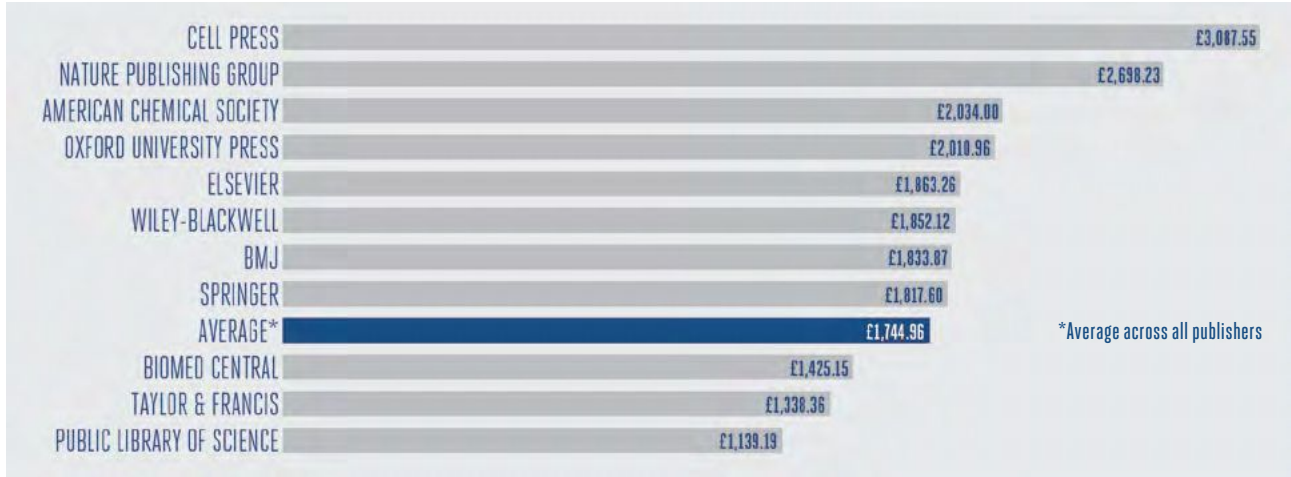
18. Björk, B. C. & Solomon, D. Op. cit.

19. Los gráficos que ilustran este artículo fueron traducidos por Visión Conjunta

20. OpenDOAR Chart - Most Frequent Languages in OpenDOAR

- Worldwide. (<http://www.opendoar.org/onechart.php?cId=&rtId=&clId=&llD=&potID=&rSoftwareName=&search=&groupby=I.IName&orderby=Tally%20DESC&charttype=bar&width=600&caption=Most%20Frequent%20Languages%20in%20OpenDOAR%20-%20Worldwide>). Consulta 12/01/2018.

21. OpenDOAR Chart - Recorded [Full-text] Data Re-use Policies - Worldwide. ([http://www.opendoar.org/onechart.php?cId=&rtId=&clId=&llD=&potID=3&rSoftwareName=&search=&groupby=pog.pogHeading&orderby=pog.pogID&charttype=pie&width=600&height=300&caption=Recorded%20\[Full-text\]%20Data%20Re-use%20Policies%20-%20Worldwide](http://www.opendoar.org/onechart.php?cId=&rtId=&clId=&llD=&potID=3&rSoftwareName=&search=&groupby=pog.pogHeading&orderby=pog.pogID&charttype=pie&width=600&height=300&caption=Recorded%20[Full-text]%20Data%20Re-use%20Policies%20-%20Worldwide)). Consulta 12/01/2018.

**FIGURA 1<sup>19</sup>. INFORME DE JISC: ARTÍCULO DE PROCESO DE CAMBIOS Y SUSCRIPCIONES**

Fuente: <https://www.jisc.ac.uk/reports/apcs-and-subscriptions>

tarse: ¿cómo contrastar con las revistas que están en acceso abierto?

Numerosos estudios reportan lo que suena conocido de manera un tanto intuitiva para el caso de Scopus y de WoS: existen áreas de mayor cobertura como las Ciencias Naturales o Ingenierías y Biomedicina en detrimento de las Ciencias Sociales, Artes y Humanidades; también es clara la preeminencia del idioma inglés en detrimento de otras lenguas<sup>12</sup>. Lo precedente también se agrava por las diferentes prácticas de difusión del conocimiento según el área del saber<sup>13</sup>. Como consecuencia de ello, la evaluación de los investigadores de distintas áreas con criterios idénticos y sobre todo más automatizados que reflexivos, determina un sesgo innegable. Más allá del *bias* mencionado, es innegable que el número de citas recibidas -reportado por ambos prestadores de bases de datos- está fuertemente correlacionado, lo que brinda un interesante nivel de confianza<sup>14</sup>; esta afirmación está vinculada naturalmente al solapamiento de títulos entre ambas bases de datos.

La base Scopus declara con acceso abierto 3.792 títulos que están con sus políticas en DOAJ<sup>15</sup>, mientras que Clarivate<sup>16</sup> anuncia unos 900 títulos con acceso abierto en la información del JCR 2015<sup>17</sup>. El número parece pequeño y las condiciones de la llamada vía dorada de las gran-

des editoriales, contra pago de APCs, no son alentadoras<sup>18</sup>; el reporte de JISC (Fig. 1) va en el mismo sentido.

Sumado a esto existe un panorama vastísimo en relación con las políticas de acceso abierto de las revistas, panorama que además no resulta muy claro a pesar de los esfuerzos de directorios como DOAJ o Sherpa Romeo.

#### LA VÍA VERDE DE LOS REPOSITARIOS

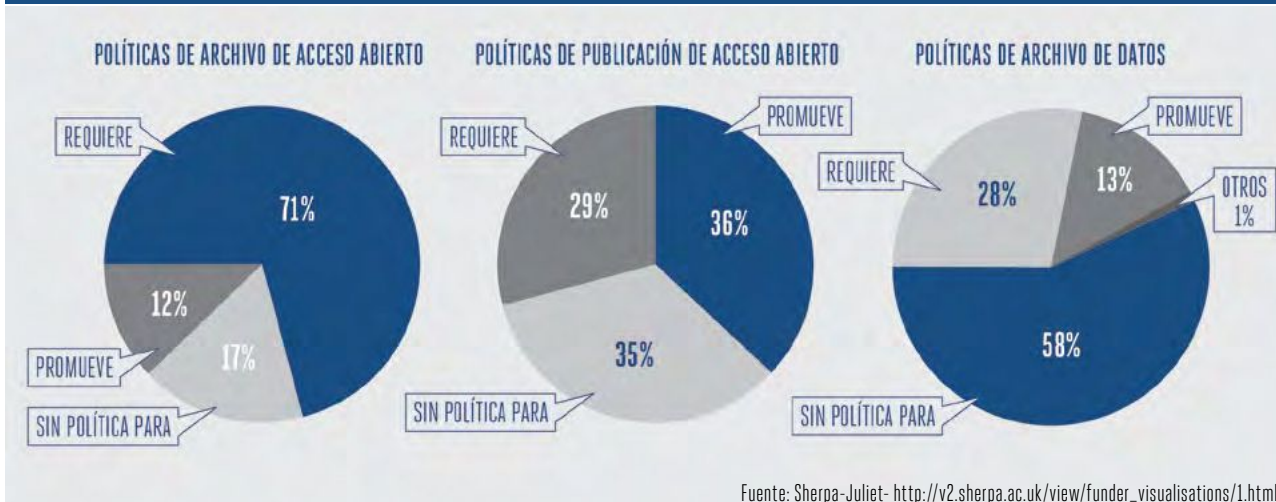
Es interesante analizar el avance de la llamada vía verde de acceso abierto, sostenida por el autoarchivo y también, por el archivo mediado de las obras en los repositorios de acceso abierto. En la consulta del 17/11/2017 el directorio OpenDOAR reportaba 3.448 repositorios en el mun-

do, claro está con diferente nivel de actualización, protagonismo y crecimiento.

El exhaustivo análisis estadístico de COAR no deja lugar a dudas de que, al menos en algunos aspectos, la situación de los repositorios reproduce el estado de la comunicación científica por la vía de las revistas. Altísima preeminencia del idioma inglés, y tan sólo un 8,9 % de los repositorios en América Latina («OpenDOAR Chart - Most Frequent Languages in OpenDOAR - Worldwide»<sup>20</sup>). La situación se agrava con el reporte de que un 74,6 % de las políticas de uso de los datos a texto completo almacenadas de los repositorios registrados no están definidas («OpenDOAR Chart - Recorded [Full-text] Data Re-use Policies - Worldwide»<sup>21</sup>) y un 74,1 %

**Hay distintos comportamientos de los propios investigadores: a la hora de actuar como lectores, pretenden acceso libre y gratuito; en cambio como autores, se inclinan por revistas de pago por suscripción o con altas APCs para lograr una evaluación positiva de sus carreras.**

**FIGURA 2. POLÍTICAS DE ACCESO ABIERTO**



Este cuadro abonaría la idea de avanzar con el acceso abierto, pero también que es necesario realizar cambios a nivel de los repositorios para que tengan un rol activo en la producción y difusión de los logros académicos y científicos.

de las políticas de contenido («OpenDOAR Chart - Recorded Content Policies - Worldwide»<sup>22</sup>) tampoco están registradas. También cabe advertir que hay preeminencia de la tipología de artículos («OpenDOAR Chart - Content Types in OpenDOAR Repositories - Worldwide»<sup>23</sup>) y una constatada dificultad para determinar el porcentaje de ellos que se encuentra accesible sin barreras a texto completo. Obviamente, en este último punto hay un problema de reproducción porque los repositorios guardan mayormente artículos que están publicados en revistas de suscripción o con APCs en las que los autores publican en principio para cumplir un circuito de evaluación tradicional que se contradice con las legislaciones nacionales de apertura de la ciencia, vigentes en distintos países. Sumado a ello, gran parte

de las políticas internacionales van por la vía verde de acuerdo a los datos de la base Sherpa Juliet (Fig. 2).

**POSIBLES SOLUCIONES**

Una visión que resulta superadora al estado actual de los repositorios institucionales es la que presenta COAR en el documento titulado “*The Next Generation Repositories*” y que sostiene la necesidad de “posicionar a los repositorios como la base para una infraestructura en red distribuida globalmente para la comunicación académica, al tope de la cual se desplegarán servicios adicionales, transformando así el sistema, haciéndolo más centrado en la investigación, abierto y solidario con la innovación, a la vez que gestionado colectivamente por la comunidad académica”<sup>24</sup>. Sin embargo, la visión precedente

no es de una concepción fácil; para dar cuenta de esto basta con desplegar los distintos niveles de intervención que se encuentran en el párrafo que la describe. El primero y más alto de tales niveles es que se habla de una infraestructura en red distribuida de manera global, es decir que las acciones a llevar adelante -particularmente los cambios estratégicos y tecnológicos-, se deben pensar en una estructura mayor que los consensúe. Esta mirada de COAR<sup>25</sup> se ve posibilitada por su propia constitución: 35 países de todos los continentes y vínculos con redes regionales como “La Referencia” en el caso de América Latina, o la Red de Repositorios de Japón por citar un caso bien diferente.

Pensar en términos globales significa, más allá de acuerdos técnicos contemplar las necesidades y los contextos (entre otros la lengua, las políticas y las prioridades) muy diversos de una manera inclusiva. Desde el punto de vista de lo tecnológico, implica seleccionar estándares internacionales y accesibles por la comunidad en su conjunto.

Para sostener esta comunidad debe existir un compromiso de las instituciones académicas y de investigación, pero está claro que ese compromiso depende de la estructura a nivel nacional

22. OpenDOAR Chart - Recorded Content Policies - Worldwide. (<http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&llD=&potID=1&rSoftWareName=&search=&groupby=pog.pogHeading&orderby=pog.pogID&charttype=pie&width=600&height=300&caption=Recorded%20Content%20Policies%20-%20Worldwide>). Consulta 12/01/2018.

23. OpenDOAR Chart - Content Types in OpenDOAR Repositories - Worldwide. (<http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&llD=&potID=&rSoftWareName=&search=&groupby=ct.ctDefinition&orderby=Tally%20ESC&charttype=bar&width=600&caption=Content%20Types%20in%20OpenDOAR%20Repositories%20-%20Worldwide>). Consulta 12/01/2018.

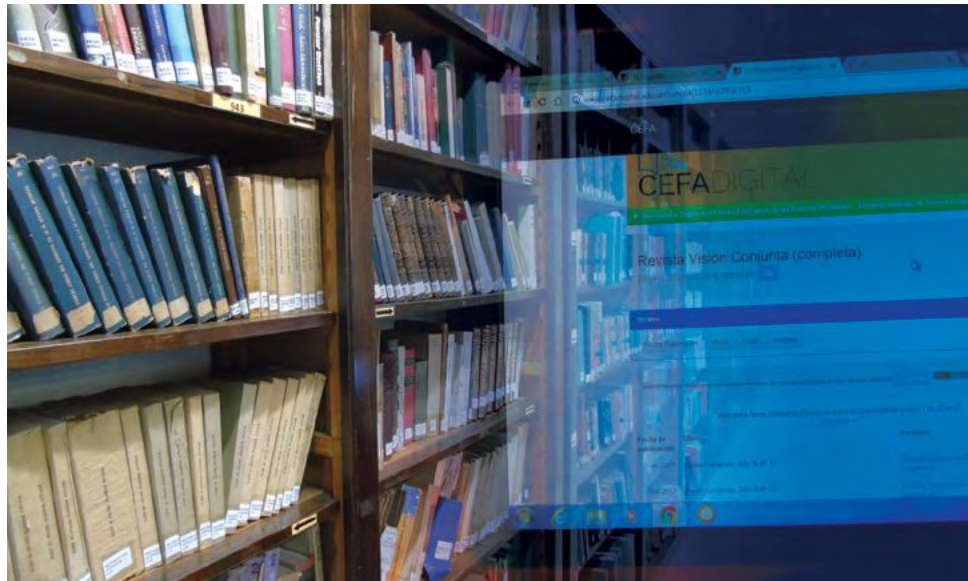
24. COAR. *Technical recommendations for next generation repositories*. (La traducción es propia).

25. Rodríguez, E., & Shearer, K. (2017). *Iniciativas y proyectos de la Confederación de Repositorios de Acceso Abierto (COAR)*. Presentado en VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'17) y XII Simposio Internacional de Biblioteca Digitales (SIBD'17) (La Plata, 2017). (<http://hdl.handle.net/10915/63552>). Consulta 16/03/2018.

que soporta al propio sistema académico científico.

Para decirlo con claridad, la comunicación académica -que en esta visión se centraría en el repositorio y naturalmente en sus protagonistas directos, es decir la institución y sus autores-, significaría la producción, evaluación y difusión de los saberes directamente en el repositorio, a la vez que una ampliación de las tipologías de esos contenidos, más allá de los artículos, tesis y documentos tradicionales, hasta incluso contribuciones menos formales o en proceso de creación. Naturalmente tal diversidad significa desafíos a la hora de analizar formatos de archivos, modos de exposición y visualización para atender por un lado a la preservación pero también al uso de esos recursos por una comunidad amplia y de saberes distintos para maximizar así el reuso de esos recursos y el beneficio no sólo para la comunidad académica sino también para la sociedad en su conjunto.

La propuesta hace de los recursos en sí mismos el eje de esta nueva generación de repositorios, por lo cual esos recursos deben estar claramente identificados, sin ambigüedad y accesibles de manera completa (no sólo sus metadatos). Los recursos deben ser portables, es decir utilizables en distintos medios, deben estar enlazados incluso con otros contenidos que los complementen o los discutan y que den lugar a servicios adicionales para la comunidad en sus diferentes estra-



tos, desde sistemas de revisión por pares, sistemas de recomendación, anotaciones sobre los recursos, estadísticas de uso y muchos otros aspectos que a la vez significarán compromisos tecnológicos para asegurar una interoperabilidad más amplia que la establecida actualmente en términos de cosecha de metadatos.

Dado que estos repositorios que vendrán tienen su eje en recursos alojados en diferentes espacios virtuales de la red propuesta, deben incluir enlaces bidireccionales que permitan su compartición bidireccional. Lo precedente asegura la ubicuidad tan deseada de los recursos y transparenta la compartición a la vez que crea nuevas oportunidades para

ampliar el alcance de la oferta de servicios brindados por los repositorios a servicios brindados por una red global de repositorios.

Un tema muy pendiente está vinculado a cuán *machine-friendly* es un recurso. Resulta muy sencillo para un operador humano distinguir en el sitio web del repositorio, por ejemplo, con qué se corresponde cada enlace; es particularmente sencillo, por citar un caso, distinguir qué enlace lleva al documento que tiene el texto completo del recurso pero esto no es sencillo para una máquina; lo mismo sucede con los derechos o los enlaces a otros recursos, por mencionar apenas un par de aspectos más. Es decir que de algún modo las nuevas tecnologías en las que se implementarán los repositorios y sus servicios, deberán brindar enlaces lo suficientemente anotados como para ser reconocidos de manera fácil por las máquinas. Está claro que pensado en términos de una red global, un sencillo y exitoso protocolo asincrónico que coseche metadatos de manera esporádica no alcanza. De algún modo tiene que existir en la red una sincronización así como alertas constantes sobre novedades de agregados o de bajas de recursos.

Parece imprescindible pensar mínimamente que los enlaces contengan una

**Los modos de implementación del acceso abierto no son sencillos y la implementación que acompaña a la legislación, muchas veces no es clara en los materiales que alcanza, las obligaciones contraídas por los investigadores y las penalizaciones en caso de no cumplimiento.**

# Pensar en términos globales significa, más allá de acuerdos técnicos contemplar las necesidades y los contextos muy diversos de una manera inclusiva. Desde el punto de vista de lo tecnológico, implica seleccionar estándares internacionales y accesibles por la comunidad en su conjunto.

anotación, una semántica estandarizada que permita la comprensión por parte de las máquinas.

En definitiva, lo que se busca es un rol más activo de los repositorios que en su implementación permita el versionado, los comentarios y el enlace entre recursos que varían en el tiempo.

Herbert van de Sompel<sup>26</sup> habla de un futuro para la publicación académica, a la que reconoce “en construcción” (Fig. 3).

### RECOMENDACIONES PARA REPOSITARIOS

En noviembre de 2017, el grupo de trabajo de COAR dedicado a establecer los comportamientos esperados y las tecnologías necesarias para la próxima generación de repositorios, incorporó en su documento una imagen provista por Petr Knoth<sup>27</sup> que distingue el estado actual de la mayoría de los repositorios y lo que se propone a futuro, remarcando la diferencia entre los servicios provis-

tos en la actualidad frente a los esperados para el futuro.

Las recomendaciones -que tuvieron su origen en las historias de usuarios y casos de uso recolectados por COAR<sup>28</sup>-, enumeran una lista de comportamientos esperados de los nuevos repositorios (11 de una lista inicial de 12 comportamientos deseados):

1. Exposición de identificadores (de los recursos).
2. Declaración de licencias (a nivel de los recursos).
3. Descubrimiento a través de la navegación.
4. Interacciones con los recursos (anotaciones, comentarios, revisiones).
5. Transferencia de recursos.
6. Descubrimiento/ posibilidad de operaciones por lote.
7. Colecta y exposición de actividades.
8. Identificación de usuarios.
9. Autenticación de usuarios.
10. Exposición de métricas de uso estandarizadas (hasta a nivel de los recursos).
11. Preservación de los recursos.

FIGURA 3. ESQUEMA DE LA FUTURA PUBLICACIÓN ACADÉMICA

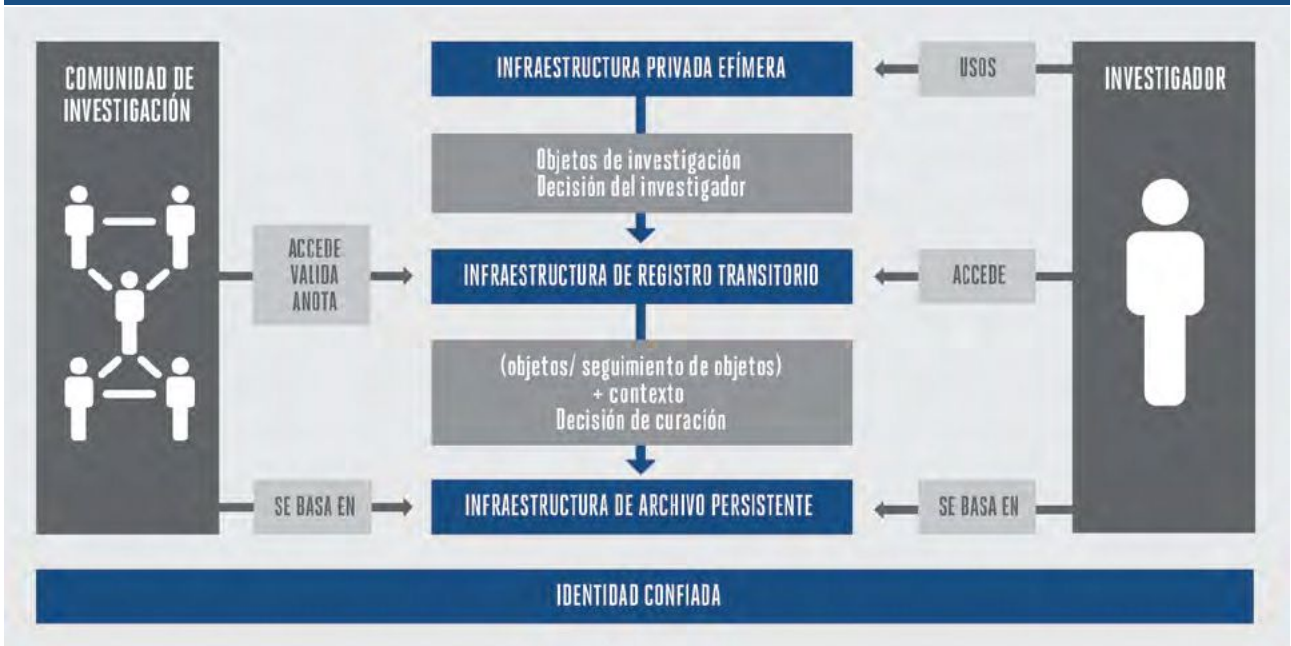


FIGURA 4. PÁGINA INICIAL DE UN ÍTEM EN SEDICI

The screenshot shows the SEDICI (Sistema de Evaluación de Documentos de Investigación Científica) interface. At the top, there is a navigation bar with 'Home', 'Search', 'Upload Material', 'Institutional', 'Frecuentes preguntas', and 'Contact'. The main content area displays the title 'Detección de bots en reportes estadísticos' by authors Catá, Juan Manuel; Lira, Ariel Jorge; and De Giusti, Marisa Raquel. The document type is 'Objeto de conferencia'. The abstract discusses the use of statistics as a tool for institutional repositories to assist in decision-making and management, highlighting the importance of precise and reliable information, especially based on access and download logs. The interface also includes search bars, social media links, and a 'Login' button.

El mencionado grupo de trabajo de COAR incluye en el documento un conjunto de herramientas recomendadas que pueden resultar adecuadas para lograr los comportamientos esperados, los que al igual que la tecnología continuarán evolucionando; el grupo ya ha abierto un repositorio en Github para compartir los desarrollos que se realicen y permitir una mayor intervención de la comunidad.

#### EJEMPLO DE ALGUNAS TECNOLOGÍAS RECOMENDADAS

Se recomienda una lectura exhaustiva del documento de COAR (“*Technical recommendations for next generation repositories*”) sobre los repositorios del futuro.

En las siguientes subsecciones, se muestran de manera muy breve algunos comportamientos deseados y las tecnologías aconsejadas.

#### PARA LA EXPOSICIÓN DE IDENTIFICADORES, LICENCIAS Y DESCUBRIMIENTO EN LA NAVEGACIÓN

En relación con la exposición de los identificadores persistentes que normalmente son una condición exigida para la integración de un repositorio a una red nacional, un usuario humano de un repositorio tras localizar un recurso, normalmente recibe también la sugerencia del URI para citar o enlazar ese recurso; aquí puede verse el enlace inicial al recurso, la sugerencia de cómo citar, al pie de la Figura 4. “Por favor, utilice este identificador (URI) para citar o enlazar este ítem: <http://hdl.handle.net/10915/57473>” que es diferente del enlace de la página de despegue (<http://sedici.unlp.edu.ar/handle/10915/57473>) y claramente, los archivos del ítem, en este caso un archivo de la ponencia y otro de la presentación,

que tendrán enlaces distintos. La operación de descubrimiento del enlace persistente que el operador humano realiza sin problemas (aunque debe copiarlo y pegarlo donde se requiera), no es posible, por ejemplo, para un *crawler*, la solución es el uso de enlaces http tipeados y la herramienta sugerida es Signposting.

Signposting es un arreglo que hace uso de encabezados para los enlaces http y de ese modo informa a un agente/máquina acerca de la naturaleza de los recursos que están en cada enlace, atendiendo a patrones que están repetidos en los portales académicos. Sirve no sólo para ubicar un localizador persistente sino una licencia, los autores de una obra o varios recursos que pueden formar parte de un ítem como es el caso de la figura precedente. Puede obtenerse mayor información en <https://signposting.org/>.

Como puede verse una misma herramienta puede ser útil para lograr los comportamientos y funcionalidades deseadas.

#### TRANSFERENCIA DE RECURSOS

La concepción central de los nuevos repositorios tiene su eje en los recursos en sí mismos, sin embargo, la idea no es hacer copias de recursos repetidos en unos y otros repositorios sino que, en la mayoría de los casos, excepto para operaciones especiales con conjuntos de recursos o problemas de latencia, lo que se precisa es referenciar dónde se encuentra el recurso. Otros casos de transferencia completa del recurso que debe contemplarse están

26. Treloar, Andrew, & van de Sompel, Herbert. *Scholarly archive-of-the-future*. Technology, 2014. (<https://www.slideshare.net/atreloar/scholarly-archiveofthefuture>). Consulta 16/03/2018.
27. Knoth, P. *Text and data mining & Recommender systems*. Presentado en COAR 2017 Annual Meeting, Venice (Italy). (<https://zenodo.org/record/580036#.WhazAVXiapo>). Consulta 16/03/2018.
28. COAR. *Technical recommendations for next generation repositories*. (<https://www.coar-repositories.org/news-media/technical-recommendations-for-next-generation-repositories/>) Consulta 12/01/2018.

#### Marisa Raquel De Giusti

Ingeniera en Telecomunicaciones, Profesor en Letras y Doctora en Ciencias Informáticas de la Universidad Nacional de La Plata. Directora de PREBI-SEDICI (UNLP) y CEGSI (CIC). Dicta clases de posgrado en la Facultad de Informática (UNLP). Investigadora Principal de la Comisión de Investigaciones de la provincia de Buenos Aires. Sus líneas de investigación incluyen el acceso abierto y la gestión de la información, los repositorios institucionales, la evaluación y la preservación de objetos digitales a largo plazo. Actúa como parte del Comité de Expertos del SNRD y dirige la iniciativa LIBLINK del Consorcio Iberoamericano para Educación en Ciencia y Tecnología (ISTEC).



vinculados, por ejemplo a operaciones que se realizan de manera externa, por ejemplo la preservación digital o la minería de datos pueden estar a cargo de otras implementaciones vinculadas al repositorio o incluso realizarse por externos.

El protocolo OAI-PMH que ha sido el estándar adoptado para la comparación de registros entre repositorios por su sencillez operativa entre otras cuestiones, supone una asincronía en estas operaciones, sin embargo cuando se precisa una sincronización mayor de

las existencias que incluya novedades sobre los recursos -nuevos, eliminados, actualizados- resulta necesario tener una lista actualizada permanentemente de URIs de recursos en el repositorio y eso supone un trabajo más sincrónico entre repositorios.

El grupo de COAR ha realizado el análisis de las tecnologías viables para lograr los comportamientos deseados y menciona a IFPS (<https://ipfs.io/>), ResourceSync (<https://ipfs.io/ResourceSync>) y SWORD. En este contexto cada repositorio (o colección) sería un posible nodo

IPFS trabajando en el contexto del protocolo ResourceSync que provee la sincronización deseada.

Mientras los *sitemaps* básicos permiten exponer un inventario de repositorio y metadatos relacionados con el rastreo, ResourceSync agrega maneras de exponer los cambios sucedidos en los recursos y proporciona una mayor expresión a los metadatos relacionados con la sincronización, así como enlaces escritos para mejorar el descubrimiento de los recursos. ResourceSync se puede usar para descubrir y sincronizar tanto el conteni-

#### GLOSARIO DE TÉRMINOS\*

- › **APCs (Article Processing Charge):** tasa de procesamiento del artículo, es decir el costo de edición de un artículo.
- › **COAR (Confederation of Open Access Repositories):** es una asociación internacional que reúne a la comunidad de repositorios y a las principales redes de repositorios para generar capacidad, alinear políticas y prácticas, y actuar como una voz global para la comunidad de repositorios.
- › **Scopus:** es una base de datos de citas y resúmenes de literatura revisada por pares: revistas científicas, libros y actas de congresos; pertenece a la editorial Elsevier.
- › **Elsevier:** editorial holandesa fundada en 1880, actualmente es una de las mayores editoriales de libros de medicina y literatura científica del mundo, que además se dedica a producir herramientas de análisis de información.
- › **WoS (Web of Science):** es una plataforma basada en tecnología Web que recoge las referencias de las principales publicaciones de disciplinas del conocimiento científico, tecnológico, humanístico y sociológico desde 1945. Es propiedad de Clarivate Analytics.
- › **Bias:** sesgo.
- › **DOAJ (Directory of Open Access Journals):** Directorio de publicaciones de acceso abierto. Es un directorio en línea que indexa y proporciona acceso a revistas de alta calidad, de acceso abierto y revisadas por pares. Todos los servicios DOAJ son gratuitos.
- › **Clarivate:** empresa que opera una colección de servicios basados en suscripción "enfocados en investigación científica y académica, análisis de patentes y estándares regulatorios, inteligencia farmacéutica y biotecnológica, protección de marcas registradas, protección de marca de dominio y gestión de propiedad intelectual". Sus servicios se enfocan principalmente en el análisis.
- › **JCR (Journal Citation Reports):** Informes sobre citas en revistas científicas. Es una publicación anual que realiza el Instituto para la Información Científica (Clarivate), que evalúa el impacto y relevancia de las principales revistas científicas del campo de las ciencias aplicadas y sociales. Permite identificar la relevancia que tiene una revista dentro de la comunidad investigadora a través de indicadores, tales como el factor de impacto (Impact factor), el índice de inmediatez (Immediacy index) o la vida media cuando es citada (cited half-life), que giran en torno a las citas recibidas.
- › **JISC (Joint Information Systems Committee):** Comité Conjunto de Sistemas de Información. Es una empresa británica cuyo objetivo es apoyar la educación y la investigación, proporcionando asesoramiento relevante y útil, recursos digitales en red, tecnología y servicios. Está financiado por una combinación de organismos de Educación Superior del Reino Unido, e instituciones de educación superior individuales.
- › **Sherpa-Romeo:** es un servicio administrado por SHERPA para mostrar las políticas de autoarchivo de derechos de autor y de acceso abierto de las publicaciones académicas. La base de datos utiliza un esquema de codificación de colores para clasificar a los editores de acuerdo con su política de autoarchivo.
- › **Open Doar (Directory of Open Access Repositories):** Directorio de Repositorios de Acceso Abierto. Es un directorio autorizado de repositorios académicos de acceso abierto. Cada repositorio ha sido visitado por el personal del proyecto para verificar la información que se registra. Brinda herramientas y soporte tanto a los administradores de repositorios como a los proveedores de servicios para compartir las mejores prácticas y mejorar la calidad de la infraestructura del repositorio. Posee buscador con variedad de criterios. Es uno de los Servicios de SHERPA, incluidos RoMED y JULIET, administrado por el Centro de Comunicaciones de Investigación (CRC). El trabajo de desarrollo es actualmente financiado por JISC, con contribuciones de la organización anfitriona del CRC, la Universidad de Nottingham.
- › **Sherpa-Juliet:** es una base de datos en línea de mandatos de acceso abierto adoptados por organismos de financiación académica. Es parte del conjunto de servicios de SHERPA en torno al acceso abierto y está dirigido por JISC.
- › **La Referencia:** red latinoamericana de repositorios de acceso abierto. A partir de los nodos nacionales se integran artículos científicos, tesis doctorales y de maestría, provenientes de más de un centenar de universidades e instituciones de investigación de los nueve países que hoy la conforman: Argentina, Brasil, Chile, Colombia, Ecuador, El Salvador, México, Perú y Costa Rica.

do como los metadatos y usa el formato XML de Sitemaps.

SWORD es un protocolo sencillo que se utiliza frecuentemente para el depósito de contenidos de un sitio a otro (<http://swordapp.org/about/>). Una aplicación concreta de este protocolo, que comparte contenidos entre un portal de revistas institucional y el repositorio puede verse en <http://hdl.handle.net/10915/27406>

## CONCLUSIONES

Los repositorios institucionales de acceso abierto han recorrido un largo y

fructífero camino y hoy no son discutidos en su relación con la difusión de los recursos institucionales e incluso por su compromiso hacia nuevas prácticas sociales y nuevas conductas de autores e instituciones en relación con la verdadera protección y resguardo de sus obras y derechos. También es reconocida la visibilidad que otorga a autores e instituciones en rankings internacionales. Sin embargo el reconocimiento de las limitaciones actuales en un horizonte tecnológico y social -que ha cambiado fuertemente desde su adveni-

miento en los albores del 2000-, obliga a cambios en prácticas y herramientas. Sólo un trabajo colaborativo entre los repositorios, sus instituciones y las redes reconocidas a nivel internacional, con acuerdos claros, uso de estándares y constante investigación y cambio, hará a los repositorios y a sus usuarios agentes activos de un cambio positivo en la esfera de la producción y compartición de conocimientos para un avance social más rápido y sin exclusiones.

## ARTÍCULO CON REFERATO

- › **Red de Repositorios de Japón:** CINI (Scholarly and Academic Information Navigator) donde se encuentra una base de datos y repositorio académico de tesis, libros y disertaciones de doctorado que se hayan publicado en distintos soportes. Está dividido en CINI ARTICLES, CINI BOOKS, y CINI DISSERTATIONS.
- › **SEDICI (Servicio de Difusión de la Creación Intelectual):** repositorio institucional de la Universidad Nacional de La Plata, un servicio libre y gratuito creado para albergar, preservar y dar visibilidad a las producciones de las Unidades Académicas de la Universidad.
- › **Bitstreams:** formato de flujo de bits es el formato de los datos encontrados en una secuencia de bits utilizados en una comunicación digital o aplicación de almacenamiento de datos. El término típicamente se refiere al formato de datos de la salida de un codificador.
- › **GitHub:** es una plataforma de desarrollo colaborativo para alojar proyectos utilizando el sistema de control de versiones Git.
- › **URI (Uniform Resource Identifier):** identificador de recursos uniforme. Es una cadena de caracteres que identifica los recursos de una red de forma unívoca. La diferencia respecto a un localizador de recursos uniforme (URL) es que estos últimos hacen referencia a recursos que, de forma general, pueden variar en el tiempo.
- › **Crawler:** Un Web Crawler (también llamado Web Spider) es un programa diseñado para explorar páginas Web en forma automática. La operación normal es que se le da al programa un grupo de direcciones iniciales, el crawler descarga estas direcciones, analiza las páginas y busca enlaces a páginas nuevas. Luego descarga estas páginas nuevas, analiza sus enlaces, y así sucesivamente.
- › **Enlaces HTTP (Hypertext Transfer Protocol):** Protocolo de transferencia de hipertexto; es un protocolo de comunicación que permite transferir información en la Web.
- › **Signposting (señalización):** herramienta que permite localizar elementos de interés a través de enlaces tipeados.
- › **Protocolo OAI-PMH (Open Archive Initiative-Protocol for Metadata Harvesting):** Protocolo de cosecha de metadatos – Iniciativa de archivos abiertos: protocolo para la transmisión eficiente de contenidos en Internet, que desarrolla y promueve estándares de interoperabilidad. Se centra en la transmisión de metadatos sobre cualquier material en soporte electrónico, y tiene como requisito esencial que éstos se codifiquen en Dublin Core.
- › **IPFS/IPFS:** protocolo hipermedia punto a punto para hacer que la web sea más rápida, más segura y más abierta. IPFS pretende reemplazar a HTTP. Cada archivo y todos los bloques dentro de él reciben una huella digital única llamada hash criptográfico. En los nodos elimina las duplicaciones en la red y rastrea el historial de versiones de cada archivo. Cada nodo de red almacena solo el contenido que le interesa y cierta información de indexación que ayuda a determinar quién está almacenando qué. Al buscar archivos, le está pidiendo a la red que busque nodos que almacenen el contenido detrás de un hash único.
- › **ResourceSync:** protocolo general de sincronización de recursos web que aprovecha XML Sitemaps. Que pueden usar los administradores de repositorios para proporcionar información que permite que los sistemas de terceros permanezcan sincronizados con los recursos en su repositorio a medida que evolucionan, es decir, se crean, actualizan y eliminan.
- › **SWORD (Simple Web-service Offering Repository Deposit):** es un estándar de interoperabilidad utilizado por los repositorios digitales para recibir depósitos de contenido desde múltiples fuentes. Esto permite enviar documentos para ser preservados y difundidos por los repositorios digitales desde herramientas externas al repositorio, como por ejemplo un sistema de gestión de publicaciones periódicas.
- › **XML (eXtensible Markup Language):** Lenguaje de Marcado Extensible: es un meta-lenguaje que permite definir lenguajes de marcas desarrollado por el World Wide Web Consortium (W3C) utilizado para almacenar datos en forma legible. Da soporte a bases de datos, es útil cuando varias aplicaciones deben comunicarse entre sí o integrar información.
- › **Sitemap:** es un archivo en el que se pueden enumerar las páginas de un sitio web para informar a Google y a otros motores de búsqueda sobre la organización de su contenido. Los rastreadores web de los motores de búsqueda, por ejemplo, el robot de Google, leen este archivo para rastrear el sitio de forma más inteligente.
- › **XML de Sitemaps:** documento describe el esquema XML para el protocolo Sitemap.