

Linked open data en bibliotecas: estado del arte

Francisco Javier Martínez Méndez, Juan-Antonio Pastor-Sánchez and Rosana López Carreño

Information Research, ISSN 1368-1613, Vol. 25, n. 2, 2020

Si bien existen desarrollos prometedores que vendrán seguidos de muchos más, la transición a las tecnologías de datos abiertos enlazados no ha sido tan rápida como se preveía. Se ha repetido varias veces que buena parte del problema tiene que ver con la curva de aprendizaje en el uso de estas tecnologías por parte de los bibliotecarios, lo que provoca cierta actitud reactiva por parte de este colectivo. También se han detectado problemas relacionados con los derechos de autor que, de momento, parecen destinar los proyectos de datos abiertos enlazados a conjuntos de datos con licencias *creative commons*. A la hora de consumir estos conjuntos de datos, los principales problemas son la coincidencia, desambiguación y alineación entre los datos de origen y los recursos de datos vinculados, junto con el mapeo del vocabulario y la reutilización de todo el conjunto publicado. La dispersión de vocabularios RDF, catálogos de conjuntos de datos, etc., complica la adopción de criterios comunes. Las bibliotecas nacionales de España, Reino Unido y Francia han desarrollado sus propias ontologías y vocabularios para publicar sus conjuntos de datos abiertos enlazados que deben dar soporte al desarrollo de nuevos proyectos. No obstante, sigue siendo complicado localizar datos del ámbito bibliotecario en los principales catálogos de conjuntos de datos. Los vocabularios SKOS están consolidados, se ve la utilidad de su publicación como datos abiertos enlazados y se percibe claramente su potencial en la reutilización de datos. Sin embargo, no sucede lo mismo con los ficheros de autoridades y los catálogos, donde son las bibliotecas nacionales las que toman la iniciativa, ya sea por su capacidad en cuanto a recursos o por el desinterés de otras instituciones (especialmente fuera de Estados Unidos). En realidad, se ha trabajado e investigado poco sobre las posibilidades reales o las necesidades de reutilización de los conjuntos de datos abiertos enlazados. Posiblemente, esta carencia de estudios influya en las decisiones a tomar por parte a nivel de dirección de las bibliotecas. Esto puede influir en que se valoren las posibilidades de esta tecnología (no porque no sea interesante, sino porque no se ha estudiado a fondo). Los conjuntos de datos abiertos enlazados en entornos bibliotecarios deben estar sujetos a la aplicación de las buenas prácticas de publicación de datos en la web. Por otro lado, los diferentes modelos de evaluación de la calidad de los metadatos en este contexto deben armonizarse siguiendo la recomendación del consorcio W3C (2017). Aunque no es objeto del presente trabajo abordar de forma exhaustiva estos aspectos de calidad, sí podemos apuntar la necesidad de un mapeado y adaptación de los diferentes modelos existentes en el marco de la definición de los requisitos y buenas prácticas de publicación de datos en la web. A pesar de todos estos

condicionantes, que sin duda alguna han lastrado la evolución hacia el uso extendido de esta tecnología, nadie parece dudar de que es el camino a seguir. Los desarrollos de los sistemas de gestión de biblioteca parecen avalar esta tesis, puesto que entre sus funcionalidades se incorporan aquellas que están relacionadas con la publicación y acceso a los registros bibliográficos como datos abiertos enlazados (especialmente, y como han destacado autores y profesionales, desde los motores de búsqueda de la web). Sin duda alguna, la consolidación del library requirements model, el estándar RDA y BIBFRAME propiciará que sus desarrolladores encuentren una base estandarizada para la representación de los objetos bibliográficos. En este punto destacan las ventajas que aportaría la consulta de catálogos bibliotecarios cuyos ficheros locales de autoridades estén vinculados con otros de gran relevancia como VIAF, ULAN, o LCNAF. Tampoco hay que olvidar las expectativas creadas alrededor de la integración de los catálogos bibliotecarios en el ecosistema de la web de datos, especialmente con iniciativas como Wikidata o con contenidos audiovisuales disponibles en la web. En definitiva, la publicación de los catálogos como conjuntos de datos RDF no se limitaría a la simple disponibilidad de los mismos en la web de datos, sino que plantearía nuevas aplicaciones y retos para estos instrumentos que irían más allá de sus usos actuales en el entorno de las bibliotecas.

<http://informationr.net/ir/25-2/paper862.html>

Conclusiones de la propia publicación

Comparación de la difusión y adopción de datos enlazados y servicios de gestión de datos de investigación en las bibliotecas

El objetivo de este estudio es identificar patrones consistentes en la difusión y adopción de innovaciones en bibliotecas a través del estudio de los datos enlazados y la gestión de datos de investigación. El servicio de gestión de datos de investigación es principalmente una innovación en el servicio, a pesar de que su implementación puede tener componentes tecnológicos. Los datos enlazados se originaron fuera de la comunidad bibliotecaria y están siendo adoptados en muchos campos. La gestión de datos de investigación emergió principalmente en bibliotecas de investigación, aunque otras instituciones, como archivos de datos, la administración de investigación de universidades y tecnologías de la información también están implicadas. Los puntos en común en la difusión y adopción de estas dos

innovaciones muy diferentes indican unos patrones consistentes que se pueden aplicar a muchas futuras innovaciones. Su difusión sigue un modelo descentralizado que se caracteriza por estar liderado por múltiples instituciones, especialmente las principales bibliotecas de investigación y algunos importantes proveedores, y por un alto grado de reinención. Esto se muestra a través de varios niveles y estrategias diversas en la adopción de estas innovaciones. Hay tres caminos identificados en la descentralización de su difusión: difusión interbibliotecaria, intrabibliotecaria y entre bibliotecarios. La adopción de ambas innovaciones es un proceso continuo que puede durar muchos años e incluir múltiples fases, entre otras: ser conscientes de su existencia, educación, preparación, adopción parcial y adopción holística. Las instituciones que adoptan estas innovaciones se pueden dividir entre pioneros y seguidores de una fase posterior, y cada una tiene comportamientos diferentes en su difusión y adopción. Se encontraron tres tipos de instituciones que facilitaron su difusión y adopción: líderes, financiadores externos y asociaciones profesionales. Se identificaron los siguientes canales de difusión: bibliografía, conferencias y reuniones, formación, redes profesionales y colaboración. Hay algunas diferencias notables en la difusión y adopción de las dos innovaciones. Respecto a la gestión de datos de investigación, las bibliotecas universitarias han liderado su implantación, mientras que en los datos enlazados se han unido algunas bibliotecas nacionales y algunos proveedores. El servicio de gestión de datos de investigación es novedoso, mientras que la adopción de datos enlazado en muchos casos significa hacer lo mismo usando tecnologías nuevas. La gestión de datos de investigación implica estudiar en profundidad las necesidades de los investigadores a través de encuestas, entrevistas, etc., mientras que para la adopción de los datos enlazados el estudio se realiza a través de proyectos experimentales que implican pruebas técnicas. Los hallazgos de este estudio revelan los retos que a los que se enfrentan las bibliotecas en la adopción de innovaciones. Una biblioteca por si sola puede no tener la infraestructura necesaria ni la experiencia para desarrollar por completo soluciones. Además, las bibliotecas no siempre pueden confiar en los proveedores para que aporten soluciones. En tercer lugar, la adopción pionera implica mucho aprendizaje e investigación. Se pretende que estos descubrimientos sean útiles para la toma de decisiones en las bibliotecas respecto a innovaciones futuras similares, en especial innovaciones en servicios y tecnologías que requieren un significativo esfuerzo intelectual.

<http://www.informationr.net/ir/25-2/paper855.html>

Resumen elaborado por Antonio Rodríguez Vela

El uso de los datos enlazados y de la Inteligencia Artificial como elementos clave en la transformación de los servicios técnicos

The Use of Linked Data and Artificial Intelligence as Key Elements in the Transformation of Technical Services

Philip E. Schreur

Cataloging & Classification Quarterly, ISSN 1544-4554, Vol. 58, n.5, 2020, p. 473-485

Los servicios técnicos de bibliotecas se han beneficiado de numerosos estímulos. Pese a que al principio se veían con suspicacia, transiciones como la que llevó de las tarjetas de catalogación a los formatos MARC han resultado ser enormemente útiles para las bibliotecas y sus usuarios. Los datos enlazados y la Inteligencia Artificial (IA) sostienen la misma promesa. A través de la conversión de los metadatos secundarios de la catalogación a datos abiertos enlazados, las bibliotecas pueden representar sus recursos en la web semántica. Pero para proporcionar alguna forma de acceso controlado a datos no estructurados, las bibliotecas deben ir más allá de la catalogación tradicional y adoptar nuevas herramientas como las que proporciona la IA para tener un acceso consistente al mundo creciente de los recursos a texto completo. Los datos enlazados ya están teniendo un impacto profundo en el proceso de los servicios técnicos. La Library of Congress ha empezado a implementar BIBFRAME, una ontología basada en RDF, para reemplazar los formatos MARC. Habrá muchos flujos de trabajo que necesitarán ser revisados para conseguir una transición exitosa al nuevo entorno, pero hay dos áreas inmediatas de preocupación: la necesidad de convertir vastas cantidades de datos heredados en datos enlazados y cómo crear y mantener autoridades en un nuevo entorno de web abierta. Al tiempo que las bibliotecas transforman este aspecto en su organización del trabajo, necesitan hacer frente a un importante número de problemas. El primero es que las autoridades tradicionales no pueden ser usadas como identificadoras para una entidad como un autor en el contexto de datos enlazados. Un segundo problema es el alto nivel de formación requerido para crear una autoridad tradicional. Pese a los retos en la transición de los metadatos bibliotecarios a los datos enlazados y a la necesidad de repensar los flujos de trabajo, está claro que esta transición supondrá enormes ventajas al armonizar los metadatos bibliotecarios con la web semántica. Se espera que se minimice la necesidad del factor humano en la creación de metadatos y su mantenimiento. Yewno es un ejemplo de una tecnología basada en IA que proporciona estrategias no tradicionales para la exploración de textos digitales. A través del uso de la lingüística computacional, la teoría del trabajo en red, las máquinas que aprenden y la IA, Yewno ayuda a los usuarios a explorar conceptos y

conexiones y a comprender materias interdisciplinarias con una interfaz gráfica. Se trata de un excelente paradigma de las posibilidades de la IA al servicio del descubrimiento del conocimiento. Al cambiar a los datos enlazados (el lenguaje de la web semántica), las bibliotecas liberan sus metadatos de los confines de un formato entendido solo por los bibliotecarios (los formatos MARC) y permite que sus metadatos se enlacen con los datos enriquecidos de la web semántica. Al expandir su visión más allá de los metadatos preasignados por los catalogadores, haciendo uso de técnicas como la IA para diferenciar conceptos y relaciones, las bibliotecas vinculan de una manera más profunda sus recursos con la web de datos. Al abrazar visiones alternativas de acceso tales como la proporcionada por Yewno, las bibliotecas pueden asociarse con los textos no estructurados y sin límites que cada vez más se añaden a su campo de acción.

<https://www.tandfonline.com/doi/full/10.1080/01639374.2020.1772434>

Resumen elaborado por Antonio Rodríguez Vela
