

EL MANEJO DE DATOS

Aproximación desde los estudios
de la información

Georgina Araceli Torres Vargas



Z666.73
L56M3

El manejo de datos. Aproximación desde los estudios de la información / Coordinadora Georgina Araceli Torres Vargas. - México: UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información, 2019.

viii, 116 pp. - Colección: TECNOLOGÍAS DE LA INFORMACIÓN.

ISBN: 978-607-30-2690-1

1. Datos vinculados. 2. Minería de datos. 3. Investigación bibliotecológica.

I. Torres Vargas, Georgina Araceli, coordinadora. II. Ser.

Diseño de portada: Natalia Cristel Gómez Cabral

Primera edición, 2020

D.R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: 978-607-30-2690-1

Publicación dictaminada

2020

Contenido

Presentación.....	7
GEORGINA ARACELI TORRES VARGAS	

MINERÍA DE TEXTO Y MINERÍA DE DATOS

Identificación de los temas de investigación en los documentos científicos del Colegio de Postgraduados.	11
ÁNGEL BRAVO VINAJA	
SANTIAGO DE JESÚS MÉNDEZ GALLEGOS	
JORGE PALACIO NUÑEZ	

Minería de texto aplicada a un diagnóstico de usuarios en Ciencia y Tecnología: aprendizajes para fortalecer la investigación bibliotecológica.	31
LOURDES FERIA BASURTO	

Minería de Datos, el caso de estudio de la Biblioteca Dr. Valentín Gómez Farías de la Facultad de Medicina de la UNAM.	43
DAVID FLORES MACÍAS	
GUADALUPE VANESA CAROLINA GUTIÉRREZ HERNÁNDEZ	

SISTEMATIZACIÓN DE DATOS Y SERVICIOS DE INFORMACIÓN

Research Data Management and Libraries: Opportunities and Challenges.....	59
KRYSZYNA K. MATUSIAK	

Integración de los principios de <i>linked data</i> en el registro bibliográfico.....	75
---	----

EDER ÁVILA BARRIENTOS

Plan para el Desarrollo de la Ciencia de Datos y Big Data (PDCDBD) en la UNAM para fines académicos y administrativos.....	95
--	----

JAVIER SALAZAR ARGONZA

Integración de los principios de *linked data* en el registro bibliográfico

EDER ÁVILA BARRIENTOS
Universidad Nacional Autónoma de México

INTRODUCCIÓN

Linked Data es un conjunto de buenas prácticas para publicar y vincular datos estructurados en el entorno de la web.

Linked Data extiende los principios de la *World Wide Web* desde la vinculación de los documentos hasta la de vincular piezas de datos y crear una Web de Datos; especifica los datos y sus respectivas relaciones, y le proporciona datos procesables por máquina a Internet. Está basado en Técnicas estándar web, pero las amplía para proporcionar el intercambio de datos y la integración. (Sakr, Sherif, *et al.* 2018, 9).

El informe sobre datos bibliotecarios enlazados del Grupo incubadora del W3C (*Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets*), es uno de los desarrollos más significativos que explica la interacción entre los principios de *Linked Data* en el ambiente de las bibliotecas, y se trata de un documento relevante para entender la oportunidad que se abre en las bibliotecas para alcanzar el impacto digital que

pretenden, además de concretar nuevos modelos en el acceso y la organización de los recursos de información.

Según Issac (2011), el Library Linked Data Incubator Group tiene como misión analizar la situación de los modelos y esquemas de metadatos y los estándares y protocolos de interoperabilidad que se deberían usar para la publicación y el uso de *Linked Data* con datos de bibliotecas.

Los subgrupos creados como parte del Library Linked Data Incubator Group, quedaron divididos de la siguiente manera:

Primer grupo. Encargado del tratamiento de datos bibliográficos de las Bibliotecas Nacionales, como la Biblioteca Británica, y las de Francia, Alemania y España.

Segundo grupo: Encargado de tratar los datos de autoridades de las mismas bibliotecas.

Tercer grupo. Encargado de conformar los diversos vocabularios controlados, la mayoría de ellos relativos a ontologías.

Cuarto grupo. Se enfoca al estudio de la presencia de recursos de información en la nube de datos enlazados (la cual se representa mediante el Linked Data Cloud Diagram).

Quinto grupo. Encargado del estudio de la relación de las citas en artículos científicos. Este estudio otorga la posibilidad de construir conjuntos de datos científicos y de mostrar sus respectivas relaciones a partir del análisis de las citas para cada artículo. Esta labor es muy interesante, ya que permite concebir nuevos e innovadores servicios de información digital y servicios documentales que apoyen el proceso de investigación, experimentación y análisis de los investigadores en diversas áreas del conocimiento.

Es importante mencionar que muchas publicaciones científicas que tienen presencia en el entorno digital utilizan metadatos para describir los datos de su publicación, lo cual facilita la creación de relaciones y en consecuencia el enlace entre datos y la creación de conjuntos o datasets.

Sexto grupo. Este grupo se encarga de darles tratamiento a los objetos digitales disponibles en el contexto de las bibliotecas. El objetivo del grupo consiste en la creación de un repositorio digital que vincule los objetos a partir del uso de metadatos semánticos.

Séptimo grupo. Se enfoca a la investigación sobre la construcción de colecciones digitales. En este grupo se abordan los alcances que FRBR puede proporcionar para la descripción de colecciones digitales. FRBR es un modelo conceptual que, al momento de ser aplicado en la organización de recursos de información, es susceptible de convertirse en un modelo descriptivo con diferentes niveles de descripción.

Octavo grupo. Aborda la presencia de la biblioteca en el entorno de las redes sociales. Un entorno muy atractivo que permite el intercambio de información entre pares y en donde los datos enlazados pueden tener una función significativa en la satisfacción de las necesidades de información de los usuarios remotos de las bibliotecas.

En suma, el trabajo de los ocho grupos se traduce en tres principales áreas de investigación:

- Área de preparación de los datos. Que se enfoca a la creación de herramientas que permitan transformar, almacenar y vincular los datos de las bibliotecas.
- Área de definición de normas. Que se encarga de abordar la construcción de normas que permitan controlar y uniformar el proceso de creación de datos enlazados.
- Área de desarrollo de interfaces. Encargada del diseño de interfaces de búsqueda y recuperación de información. La interfaz de cualquier sistema de información digital es de suma relevancia para que el usuario remoto pueda tener acceso a la información en el entorno digital.

De esta manera, se requiere contar con un método de recuperación de información que permita identificar, descubrir y acceder a los datos documentales que están disponibles en el ambiente web. Los datos de las bibliotecas que están representados en los registros bibliográficos remiten a recursos de información que pueden tener patrones de vinculación entre los datos documentales que están disponibles en diversas fuentes de la web.

Se estima que la aplicación de los principios de *Linked Data* puede favorecer el desarrollo de un método para la recuperación

de información en este ambiente. Por lo tanto, es preciso responder a tres preguntas elementales:

¿Cómo se aplican los principios de *Linked Data* en el registro bibliográfico?

¿Cómo se vinculan los datos disponibles en estos registros con otras fuentes de datos disponibles en la web?

¿Qué alcances y limitaciones tiene esta vinculación?

El propósito de este trabajo consiste en analizar la integración de los principios de *Linked Data* en el registro bibliográfico para identificar patrones de vinculación entre los datos disponibles en las bibliotecas y el ambiente web.

REVISIÓN DE LA LITERATURA

La integración de los principios de *Linked Data* en el registro bibliográfico es un tema que ha sido abordado con anterioridad. A través de la formulación teórica de la web semántica y su relación con las bibliotecas, pueden localizarse hallazgos significativos que intentan explicar dicha integración mediante estudios de caso y reflexiones teóricas, que exponen la evolución de los principios de la organización de la información y la fuerte influencia que ejercen los principios de la web semántica sobre estos postulados.

Por ejemplo, Alemu *et al.* (2012) realizaron un análisis teórico que sugiere recomendaciones para llevar a cabo un cambio conceptual de los metadatos centrados en el documento a metadatos centrados en los datos. A su vez, discutieron la importancia de ajustar los modelos de las biblioteca actuales, como RDA y FRBR a los modelos basados en los principios de los datos enlazados.

Los datos bibliográficos requieren ser tratados de manera independiente, pero inherente al recurso al cual pertenecen. Pues de esta manera el análisis de los datos bibliográficos permitirá establecer un vínculo significativo entre aquellos datos que contengan atributos similares. Para ello, es necesario que los principios bi-

bliotecológicos para la organización de la información sean compatibles con los principios de *Linked Data*.

Cole *et al.* (2013, 189) “[...] evaluaron la viabilidad y los desafíos de transformar los registros bibliográficos de bibliotecas tradicionales en *Linked Data*”. Como parte de este estudio, los autores identificaron la falta de compatibilidad entre MARC y los principios de los datos enlazados. No obstante, esta incompatibilidad ha tratado de erradicarse mediante la adopción de estrategias que tienen el propósito de integrar ambos principios. Por ejemplo, en la generación de vocabularios para la representación de los datos.

Por su parte, Tillet (2013,140) manifiesta que

[...] los datos de la biblioteca sobre nuestros recursos ya no sólo deben almacenarse en cajones de catálogo como herramienta de inventario para acceder a las colecciones de una biblioteca individual. Ahora se puede poner a disposición de cualquier persona, en cualquier lugar del mundo, en cualquier momento.

En este sentido, la disponibilidad de los datos de la biblioteca en el ambiente web permite la posibilidad de vincularlos con otras fuentes disponibles en este contexto. Siempre y cuando los registros bibliográficos se adapten al entorno actual de la web. Pues “[...] los datos enlazados ofrecen la posibilidad de realizar una reestructuración profunda del registro bibliográfico que se presenta con una nueva estructura granular” (Iacono 2014, 80).

La granularidad de los datos es una característica que les permite incrementar el nivel de detalle en su descripción y estructuración. En consecuencia, la aplicación de los principios de *Linked Data* en el registro bibliográfico beneficiará la función del registro como un medio para establecer la vinculación semántica de los datos mediante una arquitectura interoperable para la interconexión de fuentes de datos disponibles en la web.

A su vez, se ha identificado la función de *Linked Data* como método para la óptima recuperación de información en las bibliotecas. Mitchell (2016) descubrió que en los últimos dos años se

habían realizado importantes investigaciones y publicaciones que documentaban proyectos técnicos específicos, aplicaciones, vocabularios y mejores prácticas de la comunidad bibliotecaria en relación con los datos enlazados y su interacción en las bibliotecas.

La integración de *Linked Data* en el registro bibliográfico tiene dos propósitos esenciales. Por un lado, vincular los datos de las bibliotecas con otras fuentes de datos disponibles en la web. Por otra parte, propiciar la generación de un método para la óptima recuperación de la información en las bibliotecas, acorde a las demandas actuales de los usuarios. Pero los procesos de búsqueda y recuperación de la información han evolucionado, actualmente se requiere de métodos integrales que permitan el descubrimiento de información en diversos contextos relacionados.

Por lo tanto, la integración de *Linked Data* en el registro bibliográfico pone de manifiesto la interacción de normas para la descripción de los recursos, formatos de codificación y principios semánticos. RDA es la norma de descripción que en un futuro no muy lejano será aplicada en su totalidad en el ambiente de las bibliotecas. El formato MARC es el esquema de codificación por excelencia que es utilizado en los sistemas integrales de gestión de bibliotecas y es un elemento importante para la búsqueda y recuperación de recursos de información documental en las bibliotecas. La integración de RDA y MARC con los principios de *Linked Data*, ha sido motivo de análisis y discusión dentro de la literatura especializada.

Faith y Chrzanowski (2015, 133) realizaron un prototipo básico de la aplicación de RDA con RDF. Mediante los resultados obtenidos, los autores manifiestan que “[...] la vinculación de datos para una búsqueda más relevante y habilitada ayuda a abrir las bibliotecas a un mundo más amplio de posibilidades conectadas”. Además, los usuarios pueden recibir una navegación más diversa y opciones de búsqueda más sólidas a través de los datos vinculados. Los datos vinculados son un medio para conectar a más personas con información más relevante.

Por su parte, Shieh (2018) informó que el Programa de Catalogación Cooperativa de la Biblioteca del Congreso (siglas en inglés PCC-LC) ha comenzado con el mapeo de elementos entre BIBFRAME,

RDA y MARC con el propósito de mejorar de las prácticas correctas para los sistemas de recuperación de información emergentes, así como los estándares actuales de descripción actuales.

Los resultados de la búsqueda y recuperación de literatura que aborda la aplicación de los principios de *Linked Data* en el registro bibliográfico son considerables en cantidad. Sin embargo, al momento de analizar con detalle los trabajos publicados, se observa una laguna teórica y pragmática relacionada con la aplicación formal de los principios en dicho registro. La revisión de la literatura del objeto de investigación planteado en este trabajo ha permitido identificar que las breves reflexiones teóricas del objeto de estudio y las mínimas pruebas de aplicación hacen más complejo el estudio de la temática planteada. Por lo tanto, se requiere de mayor investigación teórica y metodológica para descubrir los patrones de comportamiento que se generan mediante la aplicación de los datos enlazados en el registro bibliográfico.

MANEJO DE DATOS ENLAZADOS EN LAS BIBLIOTECAS

La revolución de los datos es un fenómeno que ha sido provocado por el impacto de las tecnologías digitales en contextos donde la información es un elemento trascendental para la generación de nuevos conocimientos. En la actualidad, los datos son producidos a gran velocidad y de manera continua por personas, computadoras y como parte del uso de aplicaciones comerciales y de geolocalización.

Smith (2014) ha explicado con anterioridad la utilización de métodos y tecnologías digitales para el procesamiento de los datos disponibles en las bibliotecas. Uno de los aspectos mas sobresalientes del procesamiento de los datos, recae en la cercana relación que tiene con los procesos de descripción y catalogación de los recursos.

Se puede afirmar que la catalogación y descripción de los recursos son métodos analíticos e intelectuales para el procesamiento y la obtención de datos bibliográficos. Las bibliotecas contienen enormes cantidades de datos de índole bibliográfica y documental que son registrados con formatos altamente especializados. Los

datos de la biblioteca son generados como parte de procesos intelectuales mediante el uso de normas, vocabularios y principios de índole bibliotecológica.

Los datos de la biblioteca son utilizados como puntos de acceso para los recursos de información documental que son registrados en los catálogos. Estos datos son almacenados y registrados en herramientas para la búsqueda y recuperación de información. Por lo tanto, los registros bibliográficos de las bibliotecas se convierten en una fuente para la construcción de datos enlazados en el ambiente de las bibliotecas.

El manejo de datos enlazados pone de manifiesto un método para la eficiente publicación y vinculación de datos estructurados en la web. Los datos que pertenecen a los registros bibliográficos de las bibliotecas son de gran utilidad para la generación de datos enlazados de tipo bibliográfico y de autoridad.

Los registros bibliográficos codificados en formato MARC deben adaptarse a los principios de *Linked Data*. el proceso de adaptación de estos registros es sistematizado y se encuentra relacionado con el uso de normas y estándares de índole internacional. La desfragmentación del registro bibliográfico dará como resultado el tratamiento individual de los datos que caracterizan a los recursos de información documental.

Cada dato bibliográfico puede vincularse semánticamente con otros datos de atributos similares y que estén disponibles en el ambiente web. Para llevar a cabo esta vinculación es necesario ejercer buenas prácticas de manejo de datos enlazados en las bibliotecas. El proceso para el eficiente manejo de datos enlazados se compone de las siguientes fases:

- A. Fase de selección de datos. Los registros de datos bibliográficos que serán seleccionados pueden pertenecer a una determinada colección. Formar parte de una temática en un dominio específico de conocimientos. La especificidad de los datos seleccionados es un asunto importante, pues a mayor especificidad en los datos será mayor el grado de exactitud que se alcance al momento de vincularlos.

Una vinculación semántica de datos es un procedimiento intelectual que además de conectar a los datos, tiene el propósito de explicar el significado de la relación que se establece entre ellos.

- B. Fase de normalización de los datos. La estructuración de los datos es un procedimiento normalizado. Para ello se utilizan los principios de *Linked Data* señalados por el W3C. La asignación de URIs a los datos, la codificación de los datos con RDF y la utilización de vocabularios estandarizados son algunas de las acciones que se desarrollan en esta fase. Además, se debe contemplar el uso de los principios bibliotecológicos y los lenguajes documentales que favorezcan la estructuración de los datos y beneficien su descripción y acceso.
- C. Fase de descripción de los datos. El registro de los datos debe ser exacto, sin ambigüedades y libre de inexactitudes. RDA y Dublin Core, proporcionan elementos descriptivos para representar los atributos bibliográficos y de contenido de los recursos. Los tesauros, folksonomías y ontologías pueden favorecer la descripción temática de los recursos. La óptima descripción temática del recurso permitirá obtener datos de mayor precisión para establecer vinculaciones de mayor significado entre los datos.
- D. Fase de vinculación de los datos. Los datos debidamente descritos y estructurados deberán vincularse entre sí. Las vinculaciones de los datos deben explicar la relación que existe entre ellos y el significado que los rodea en un determinado contexto. Pues no se trata únicamente de establecer conexión entre los datos, sino de explicar el significado de la vinculación que se establece entre los datos disponibles en diversas fuentes. Para vincular los datos es necesario establecer interoperabilidad entre las fuentes que serán conectadas.
- E. Fase de recuperación y acceso a los datos. Los datos enlazados generados en las bibliotecas deben ser abiertos, libres de cualquier restricción técnica, legal y económica. Para

ello, deberán aplicarse los principios de licencias abiertas de datos. La recuperación de los datos enlazados deberá permitir el descubrimiento de los datos y de sus respectivas vinculaciones. Además de la tradicional búsqueda textual, será necesario visualizar gráficamente las vinculaciones mediante el uso de una interfaz para la consulta de grafos.

- F. Fase de preservación de los datos. La prospectiva del uso de los datos debe sujetarse a un proceso planificado. Se debe contemplar qué datos será necesario conservar para ser utilizados en un futuro y garantizar su acceso sin que importen los rápidos cambios tecnológicos del contexto que los rodea. Es deseable contar con una política de preservación de datos que respalde el proyecto de datos enlazados y su generación en las bibliotecas.

El óptimo manejo de los datos enlazados en las bibliotecas pone de manifiesto la figura de un cambio de paradigma relacionado con la catalogación de los recursos de información. Pues se estima que la aplicación de *Linked Data* en el ámbito de la organización de la información, da la pauta para la descripción semántica de los recursos; en el siguiente apartado se abordan algunos de los principios identificados que explican la formulación de este proceso.

HACIA LA DESCRIPCIÓN SEMÁNTICA DE LOS RECURSOS

RDA (2014) menciona tres tipos de descripciones de recursos:

- A. Descripción compresiva. Se utiliza para describir a los recursos como un todo. Se emplea para describir cualquier tipo de recurso.
- B. Descripción analítica. Es utilizada para describir una parte de un recurso más amplio.
- C. Descripción jerárquica. Une los dos tipos anteriores, es decir, combina una descripción integral de un recurso, con la descripción analítica de una o más de sus partes, por ejem-

plo: se describe un libro y además un capítulo del mismo; una revista científica y sus respectivos artículos.

La descripción semántica del recurso reúne los tipos de descripciones anteriores y explica la vinculación de los datos que son descritos y que pertenecen a los registros. Pues no se trata únicamente de establecer relaciones superficiales entre los recursos. Sino de explicar la conexión y darle significado a la vinculación que se establece entre los datos de los recursos.

Por lo tanto, la descripción semántica de los recursos se define como un proceso intelectual y apegado a normas que tiene el propósito de registrar y representar los atributos bibliográficos, temáticos y de autoridad de los recursos para explicar el significado de las vinculaciones existentes entre los datos pertenecientes a estos recursos.

Indudablemente, los métodos y estándares para organizar la información han cambiado, y se han adaptado a las características de los diferentes tipos de recursos que han surgido y han incorporado el uso de las tecnologías digitales para su consulta. Sin embargo, los principios sustanciales en los que descansa la organización y recuperación de la información continúan siendo los mismos. Localizar la información, identificar si es la información que necesitamos para tomar la decisión de obtenerla. En estos principios descansan desde el desarrollo de los catálogos hasta el de la web semántica (Martínez 2009, 12).

Los principios sustanciales de la organización de la información deben evolucionar de acuerdo con las exigencias del contexto de información actual. Hoy en día los datos de las bibliotecas están cobrando mayor relevancia para satisfacer las demandas informativas de la comunidad.

Actualmente los datos de la biblioteca remiten a una amplia gama de recursos de información documental. Algunos de estos recursos se vinculan de manera directa con otros contextos fuera de la biblioteca, por ejemplo, con bases de datos, repositorios y plataformas de contenidos digitales. Es necesario definir métodos que permitan identificar estas conexiones de una manera automatizada y accesible para el usuario final.

RESULTADOS DE LA INTEGRACIÓN DE LINKED DATA EN EL REGISTRO BIBLIOGRÁFICO

Linked Data reúne los componentes principales para desarrollar la web semántica. Berners-Lee (2006), definió cuatro reglas básicas para construir datos enlazados:

1. Utilizar URIs para nombrar a las cosas disponibles en la web.
2. Utilizar el protocolo HTTP-URI para que los usuarios de la web puedan buscar esas cosas.
3. Cuando un usuario busca un URI, debe proporcionar información útil empleando los estándares RDF y SPARQL.
4. Incluir enlaces a otros URIs para que el usuario pueda descubrir más cosas.

Linked Data es un concepto de propósito general. Literalmente cualquier cosa puede ser descrita utilizando datos enlazados.

RDF proporciona un modelo común para datos enlazados y es particularmente adecuado para representar datos en la Web. *Linked Data* utiliza RDF como su modelo de datos y lo representa en una de varias sintaxis (Wood, Zaidman & Luke 2014, 9).

Candela, *et al.* (2015) llevaron a cabo un prototipo de implementación de *Linked Data* en un contexto de datos bibliográficos. Migraron 200 mil registros del catálogo de la Biblioteca Miguel de Cervantes a una nueva base de datos relacional cuyo modelo de datos se adhiere a las especificaciones FRBR y FRAD. El contenido de la base de datos se asignó posteriormente a tripletas RDF que emplean el vocabulario de RDA para describir las entidades, así como sus propiedades y relaciones.

A su vez, Possemato (2018), realizó una investigación en donde se ocupa de la aplicación del estándar RDA en el campo de los datos vinculados y de cómo se puede utilizar este estándar para mejorar la calidad de los datos producidos por las bibliotecas y al-

canzar así las ventajas que la web semántica puede aportar al sector del patrimonio cultural.

Las dos investigaciones anteriores, fueron relevantes para comprender el camino a seguir en la implementación de *Linked Data* en el registro bibliográfico. Pues a pesar de haber localizado una considerable muestra de estudios de caso, la mayoría de ellos eran abordados desde una perspectiva informática o enfocada al ámbito computacional y tomaban como objeto de estudio datos de diversa tipología y naturaleza disciplinar.

De esta manera, para integrar los principios de *Linked Data* en el registro bibliográfico, se tomó en cuenta la obra de *El nombre de la rosa*, escrita en el año 1980 por el filósofo italiano Humberto Eco. Se utilizaron los elementos núcleo para la descripción de manifestaciones señaladas en RDA y el formato MARC para obtener dos ejemplificaciones básicas de la construcción del registro.

Se aplicaron los principios básicos de RDF para obtener la estructura general del recurso que permitieran representar a los datos bibliográficos como un triple. A cada dato del registro bibliográfico le fue asignado un URI. Cada dato bibliográfico debe contar con un URI único e individual que tenga dependencia directa con el sistema que los almacena y genera. En este sentido, cada URI es irrepetible y representa a un dato dentro de un dominio específico. Cada URI debe estar normalizado bajo principios interoperables que le permitan vincularse con otra fuente ajena a la biblioteca. La interoperabilidad entre los datos permitirá consultarlos en diferentes plataformas y dispositivos.

Ilustración 1. Ejemplo de registro MARC con elementos estructurales de Linked Data.
Fuente: elaboración propia, 2018.

SUJETO URI: http://el_nombre_de_la_rosa_work_rda	
PREDICADO	OBJETO
Elemento MARC	Registro de los datos
http://marc.008/35-37	http://spa
http://marc.020	\$a http://ISBN_978-970-810-026-7
http://marc.040	\$a http://rda
http://marc.100	\$a http://Eco_Umberto \$d http://1932-2016 \$e http://Autor
http://marc.245	\$a http://El_nombre_de_la_rosa \$c http://Umberto_Eco
http://marc.250	\$a http://Segunda_edici3n
http://marc.264	\$a http://M3xico \$b http://Random_House_Mondadori \$c http://2004 \$c http://1980
http://marc.300	\$a http://783_p3ginas
http://marc.336	\$a http://Texto
http://marc.338	\$a http://Volumen
http://marc.650	http://Novela_hist3rica
http://marc.700	\$a http://Ricardo_Pochtar \$e http://Traductor

Los datos bibliográficos deben ser desfragmentados para luego ser tratados semánticamente bajo los principios de *Linked Data* pero, se deben eliminar puntuaciones, pues éstas son irrelevantes para la búsqueda, recuperación y vinculación de los datos. En el caso del formato MARC, los subcampos e indicadores de codificación, pueden llegar a alterar la representación de los datos en formato RDF. Sin embargo, es posible utilizar los datos que están colocados en las diversas áreas del registro bibliográfico codificados en formato MARC.

Ilustración 2. Ejemplo de registro RDA con elementos estructurales de *Linked Data*.
Fuente: elaboración propia, 2018.

SUJETO URI: http://el_nombre_de_la_rosa_work_rda	
PREDICADO	OBJETO
Elemento RDA	Registro de los datos
http://rda.título	http://El_nombre_de_la_rosa
http://rda.mención_de_responsabilidad	http://Umberto_Eco
http://rda.edición	http://Segunda_edición
http://rda.lugar_de_publicación	http://México
http://rda.editor	http://Random_House_Mondadori
http://rda.fecha_de_publicación	http://2004
http://rda.copyright	http://1980
http://rda.identificador_de_la_manifestación	http://ISBN_978-970-810-026-7
http://rda.sopORTE	http://Volumen
http://rda.extensión	http://783_páginas
http://rda.tipo_de_contenido	http://Texto
http://rda.idioma_de_la_expresión	http://Español
http://rda.creador	http://Eco_Umberto_1932-2016
http://rda.colaborador	http://Ricardo_Pochtar
http://rda.designador_de_relación	http://Traductor
http://rda.relación_temática	http://Novela_histórica

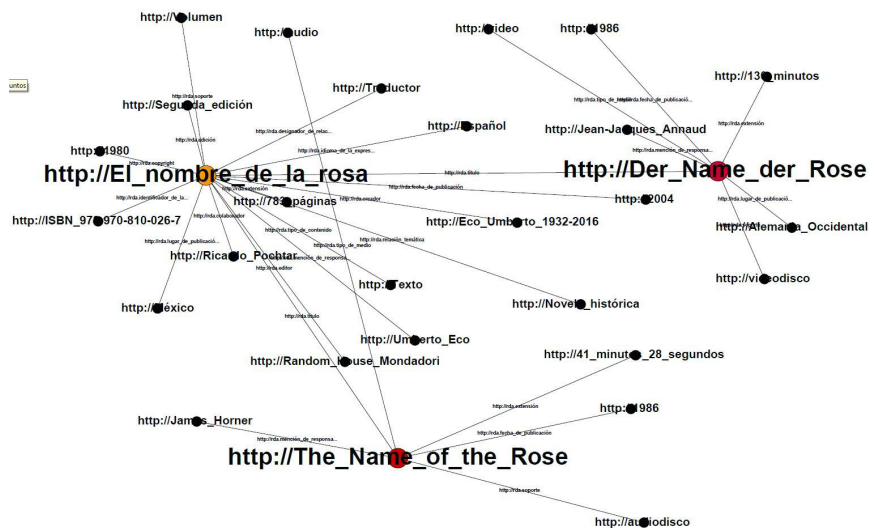
Los elementos de RDA presentan mayor flexibilidad en su integración con los principios de *Linked Data*. Sin embargo, también es necesario omitir la puntuación al momento de registrar los datos. RDA plantea la posibilidad de establecer relaciones bibliográficas entre los recursos. Los designadores de relación son elementos que fomentan el establecimiento de conexiones entre los datos. Sin embargo, se trata de una relación meramente superficial que carece de una explicación semántica.

Cuando los datos del registro bibliográfico son desfragmentados y tratados de manera individual, es posible construir grafos

RDF que ayudan a comprender su comportamiento en un determinado contexto o dominio. Según Powell y Hopkins (2015, 9), “un grafo es un modelo abstracto conceptual que sirve para representar algunos aspectos del mundo que se pueden detectar, observar o inferir.” Un grafo puede utilizarse como un método visual para comprender las relaciones que se establecen entre los datos que pertenecen a una realidad específica.

En este sentido, los grafos RDF son desarrollados siguiendo una lógica entre sujeto, predicado y objeto. En la *Ilustración 3*, puede observarse la construcción de un grafo RDF que representa a los datos de los registros RDA y MARC de la obra “El nombre de la rosa”. En color naranja, se puede visualizar el nodo correspondiente al título de la obra. El cual se trata de una versión impresa del libro (manifestación RDA). Los nodos rojos representan la videograbación (*Der name der rose*) y grabación sonora (*The name of the rose*) que se desprenden de la obra principal.

Ilustración 3. Grafo de la vinculación de datos bibliográficos disponibles en los registros MARC y RDA. Fuente: elaboración propia, 2018.



Los datos de color azul corresponden a los elementos RDA que tienen la función de predicado en el grafo. Cada elemento RDA explica el significado que tiene la vinculación efectuada entre los datos que están sombreados de color verde. Cada uno de los datos representados en el grafo tiene la capacidad de vincularse con atributos similares. La explicación del significado de la vinculación se expresa en el grafo mediante una visualización integral.

La consulta de los datos enlazados necesita de una visualización grafica que contribuya a la comprensión del comportamiento de los datos que son vinculados en un determinado contexto. En el ámbito de las bibliotecas, los datos bibliográficos se encuentran en constante movimiento, ya sea mediante actualizaciones que sufren o con la generación de nuevos recursos que permiten desarrollar redes de datos bibliográficos más extensas y complejas.

En la actualidad los datos bibliográficos deben contemplarse como elementos que sirven para desarrollar estructuras complejas que pueden vincularse con otros contextos independientes a las bibliotecas. El potencial de los datos bibliográficos dependerá de su capacidad para vincularse en otros contextos pertinentes y significativos que se caractericen por contar con información arbitrada y de calidad. Adicionalmente, los datos deben ser publicados de manera abierta, sin barreras de índole técnica, legal y económica.

CONSIDERACIONES FINALES

La integración de los principios de *Linked Data* en el registro bibliográfico, es un proceso intelectual y fundamentado basado en el uso de normas como RDF, URIs y SPARQL. Se requiere que los datos bibliográficos tengan mayor flexibilidad para adaptarse a entornos interoperables de datos abiertos enlazados. Para ello, es necesario que el registro bibliográfico pueda vincularse con fuentes externas al catálogo de la biblioteca.

Mediante la aplicación de los principios de *Linked Data* en el registro bibliográfico, fue posible identificar la necesaria flexibilidad que requiere RDA y MARC para adaptarse a los principios de

los datos enlazados. Si bien es posible asignar URIs a cada uno de los datos del registro y estructurarlos bajo RDF, es necesario que los subcampos del formato MARC tengan mayor adaptabilidad con los principios de *Linked Data*.

Los elementos de RDA manifiestan una mayor flexibilidad de adaptación a los principios de *Linked Data*. Sin embargo, se requieren ejercicios de integración con mayor grado de complejidad y detalle que permitan identificar el comportamiento de los datos bibliográficos mediante su estructuración semántica. Proyectos como *RDA Registry* han generado vocabularios RDA-RDF que será necesario abordar en futuras investigaciones mediante la generación de prototipos de registro de datos enlazados de índole bibliográfica.

El grafo RDF es un método para visualizar los datos enlazados generados. El acceso y consulta de los datos enlazados pone de manifiesto la generación de este tipo de grafos. Así la consulta de los datos y sus respectivas vinculaciones serán más usables y accesibles mediante este método gráfico.

Se estima que la generación del BIBFRAME por parte de LC y LRM de IFLA, fomenten una mayor adaptabilidad, interoperabilidad y flexibilidad de los datos disponibles en el registro bibliográfico con los principios de *Linked Data*. Sin embargo, será necesario analizar los alcances y limitaciones de ambos desarrollos para conformar entornos de datos enlazados bibliográficos.

OBRAS CONSULTADAS

Alemu, Getaneh, Brett Stevens, Penny Ross, and Jane Chandler. 2012. "Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-Based Data Models." *New Library World* 113 (11): 549–70. DOI:10.1108/03074801211282920.

Berners-Lee, Tim. "Linked data, publicada el 27 de julio de 2006", <https://www.w3.org/designissues/linkedata.html>

- Candela, G., P., Escobar, M., Marco-Such, R.C., Carrasco. 2015. "Transformation of a library catalogue into RDA linked open data." In *lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bio-informatics)*, 9316: 5–7. DOI:10.1007/978-3-319-24592-8.
- Cole, Timothy W., Myung Ja Han, William Fletcher Weathers, and Eric Joyner. 2013. "Library MARC Records Into Linked Open Data: Challenges and Opportunities." *Journal of Library Metadata* 13 (2–3): 163–96. DOI:10.1080/19386389.2013.826074.
- Faith, Ashleigh, and Michelle Chrzanowski. 2015. "Connecting RDA and RDF: Linked Data for a Wide World of Connected Possibilities." *Pennsylvania Libraries: Research & Practice* 3 (2): 122–35. DOI:10.5195/PALRAP.2015.106.
- Iacono, Antonella. 2014. "Dal Record Al Dato. Linked Data e Ricerca Dell'informazione Nell'OPAC." *Italian Journal of Library, Archives, and Information Science*, 5 (1): 77–102. DOI:10.4403/jlis.it-9095.
- ISAAC, Antoine, *et al.* 2011. "Library Linked Data Incubator Group: Datasets, Value, Vocabularies, and Metadata Element Sets: W3C Incubator Group Report", <http://www.w3.org/2005/Incubator/llld/XGR-llld-vocabdataset-20111025/>
- Joint Steering Committee for Development of RDA. 2014. *Resource Description & Access*. American Library Association, Canadian Library Association, CILIP.
- Martínez Arellano, Filiberto Felipe. 2009. "Organización de la información: del catálogo a la web semántica". En *Memoria del XXVI Coloquio de Investigación Bibliotecológica y sobre la Información*, 1, 2 y 3 de octubre de 2008 compiladores Filiberto Felipe Martínez Arellano, Juan José Calva González, 3–14. México: UNAM. Centro Universitario de Investigaciones Bibliotecológicas.
- Mitchell, Erick. 2016. "Library Linked Data: Early Activity and Development." *Library Technology Reports*. Vol. 52. DOI:10.5860/ltr.52n1.

- Possemato, Tiziana. 2018. "How RDA Is Essential in the Reconciliation and Conversion Processes for Quality Linked Data." *Italian Journal of Library, Archives, and Information Science*, 9 (1): 49–61. DOI:10.4403/jlis.it-12447.
- Powell, James & Matthew, Hopkins. 2015. *A librarian's guide to graphs, data and the semantic web*. USA: Elsevier.
- Sakr, Sherif, Marcin Wylot, Raghava Mutharaju, Danh Le Phuoc, and Irini Fundulaki. 2018. Linked Data: Storing, Querying, and Reasoning. *Linked Data: Storing, Querying, and Reasoning*. USA: Springer Link. DOI:10.1007/978-3-319-73515-3.
- Shieh, Jackie. 2018. "Reports from the Program for Cooperative Cataloging Task Groups on URIs in MARC & BIBFRAME." *Italian Journal of Library, Archives, and Information Science*, 9 (1): 111–20. DOI:10.4403/jlis.it-12429.
- Smith, K.M. (2014). *Handbook of data processing for libraries: modern methods and latest technologies*. London: Koros Press.
- Tillett, Barbara. 2013. "RDA and the Semantic Web, Linked Data Environment." *Italian Journal of Library & Information Science* 4 (1): 139–45. DOI:10.4403/jlis.it-6303.
- Wood, David, Marsha, Zaidman y Ruth, Luke. 2014. *Linked data: structured data on the web*. Estados Unidos de América: Manning.

Manejo de datos. Una aproximación desde los estudios de la información. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Francisco Xavier González y Ortiz; revisión de pruebas, Valeria Guzmán González; formación editorial, Natalia Cristel Gómez Cabral. Instituto de Investigaciones Bibliotecológicas y de la Información / UNAM. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en febrero de 2020.